

From: Li, Ruixiang
Sent: Tuesday, October 30, 2001 2:41 PM
To: STIC-Biotech/ChemLib
Subject: Sequence search of Application NO: 09/775,181

Please do a standard search of SEQ ID NOS: 1, 2, and 4 against both the commercial and interference nucleic acid databases.

Thank you very much!

Ruixiang Li
GAU 1646
CM1 8E18
Mail Box 10C01
306-0282

*Edward Hart
Technical Info Specialist
STIC / Biotech
CM1 12C14 Tel: 305-9203*

SEARCH REQUEST FORM

Scientific and Technical Information Center

Access DB#

53994

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30: _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc. if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: _____	NA Sequence (#) <u>1</u> STN	
Searcher Phone #: _____	AA Sequence (#) <u>2</u> <u>Reverse</u> <u>ONTA</u>	
Searcher Location: _____	Structure (#) _____	Questel Orbit _____
Date Searcher Present: <u>10/30/01</u>	Bibliographic _____	Dr. Linn _____
Date Timed out: <u>11/5/01</u>	Litigation _____	Lexis Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems <u>04</u>
Client Prep Time: _____	Patent Family _____	WWW Internet _____
File No Time: _____	Other _____	Other Specialty _____

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2001, 06:44:17; Search time 3196.64 Seconds
(without alignments)
17651.768 Million cell updates/sec

Title: US-09-775-181-1
Perfect score: 3648
Sequence: 1 atgggagccatggcttacc.....gggatatgttttaaaagtgtag 3648

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1344157 seqs, 7733874588 residues 2688314
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_em:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pl1:*
- 13: gb_pl2:*
- 14: gb_pl3:*
- 15: gb_pl4:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rod:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_hum1:*
- 34: em_hum2:*
- 35: em_hum3:*
- 36: em_hum4:*
- 37: em_hum5:*
- 38: em_hum6:*
- 39: em_hum7:*
- 40: em_in:*
- 41: em_in:*
- 42: em_or:*
- 43: em_or:*

- 44: em_ov:*
- 45: em_pat:*
- 46: em_ph:*
- 47: em_pl:*
- 48: em_ro:*
- 49: em_sts:*
- 50: em_sy:*
- 51: em_un:*
- 52: em_vi:*
- 53: gb_sts1:*
- 54: gb_sts2:*
- 55: gb_sts3:*
- 56: gb_sy:*
- 57: gb_un:*
- 58: gb_vil:*
- 59: gb_vil2:*
- 60: gb_vil3:*
- 61: gb_vil4:*
- 62: gb_vil5:*
- 63: gb_vil6:*
- 64: gb_vil7:*
- 65: gb_vil8:*
- 66: gb_vil9:*
- 67: gb_vil10:*
- 68: gb_vil11:*
- 69: gb_vil12:*
- 70: gb_vil13:*
- 71: gb_vil14:*
- 72: gb_vil15:*
- 73: gb_vil16:*
- 74: gb_vil17:*
- 75: gb_vil18:*
- 76: gb_vil19:*
- 77: gb_vil20:*
- 78: gb_vil21:*
- 79: gb_vil22:*
- 80: gb_vil23:*
- 81: gb_vil24:*
- 82: gb_vil25:*
- 83: gb_vil26:*
- 84: gb_vil27:*
- 85: gb_vil28:*
- 86: gb_vil29:*
- 87: gb_vil30:*
- 88: gb_vil31:*
- 89: gb_vil32:*
- 90: gb_vil33:*
- 91: gb_vil34:*
- 92: gb_vil35:*
- 93: gb_vil36:*
- 94: gb_vil37:*
- 95: gb_vil38:*
- 96: gb_vil39:*
- 97: gb_vil40:*
- 98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1634.4	44.8	4595	85	AB032962 Homo sapi
2	1304.4	41.2	151218	79	AL355587 Homo sapi
c	1443	39.6	185146	79	AL355542 Homo sapi
4	1077.8	29.5	1857	85	AB052146 Macaca fa
5	903	24.8	11117	80	AL358932 Homo sapi
5	903	24.8	19526	79	AL139821 Homo sapi
c	227.2	6.2	143878	80	AL359974 Homo sapi
c	225.6	6.2	159596	79	AL161654 Homo sapi

Db 601 GTGGTGGCCGTTGGTGTGAAGTGCAGCAAGCGCTCAACCTCAGCTCAGAGAAGAAAC 660
Qy 2673 tgggacccacgaacatcgatgttgcagaaagtctctcagtgctcatagcaagcccaagga 2732
Db 661 TGGCACCACGACATCGATGTTACAGAAAGTCTCTCAGTGTCTATAGCAAGCCCAAGGA 720
Qy 2733 gaagactcttgattagctgggaaaccccaacagcagggtgtggaagaacgcactaaac 2792
Db 721 GAAGACTCTTGGATTAGCTGGGAAACCCCAACAGCAGAGGTGTGGAAGACGCATTAATC 780
Qy 2793 ccagaaacctttccaaaagataaagagacaaacagaaatcaactcaattctgataaac 2852
Db 781 CCAGAAACCTTTCCAAAGATAAAGACAAACAGAAATCACTCAAAATTTCTGATAACAC 840
Qy 2853 agagactaaagatctctgcccccaaaactcaaatctctgcagagagcgaagaagcctca 2912
Db 841 AGAGACTAAAGATCTCTGCCCCCAAACTCAATCTTGCAGGAGGCCAGAAAGCCTCA 900
2913 gaaatctgggattatgaacacaaaggggtcaacccccaccactgccaatctgacctgaa 2972
901 GAAATCTGGGATTATGAACAACAAGGGTCAACCCACCCTGCCAATTTGACCTGAA 960
Qy 2973 ccaggcaccacccagatgaagacaaactttgacattggggaggtgtgtccttgggaagt 3032
Db 961 CCAGGCACCACCAGATGAAGACAACTTTGACATTTGGGGAGGTGTCTTTGGGAGGT 1020
Qy 3033 ttatgacctgacctgtgctgtgcttcagatcaataaaagtccaagcagcagtatctat 3092
Db 1021 TTATGACCTGACCCCTGCTCTCTGCTTCAGATCAAAAGTTCAAAAGCAGCTATCTAT 1080
Qy 3093 tgtgctcttgaaatggagaaaaacccccacttttccctaaaggagaaatctcaccacaa 3152
Db 1081 TGTGGCTTCTGAAATGAGAAAAACCCCACTTTTCTTAAAGGAGAAATCTCACCACAA 1140
Qy 3153 gctaagcgactgaggttctgagcaatccaatcagaacgcagatagatgaagctgaagt 3212
Db 1141 GCCTAAGGCGAGCTGAGGTTTGTGAGCAATCCAAATCAGAGCGCATAGATAAGCGTGAAGT 1200
Qy 3213 atgccttggagagcgaagcagtcctatttggaaagtgaagaagcttttatttccaa 3272
Db 1201 ATGCTTTGGGAGAGCCAGCCAGTCCTCAATTTTGGAGATGAGAGCTTTTGATTCCAA 1260
Qy 3273 gactccagttctccagagagggcaaaagagggaacgagggttcagcctcgtgcagccaa 3332
Db 1261 GACTCCAGTTCTCCACAGAGGGCAAAAGAGAGAGAGCGAGGTGAGCTGTCAGGCCAA 1320
3333 tgtgtgctggcagagcgaagaactgcccccccaagctgtagcatcaaaaacagagaa 3392
1321 TGTGTGCTGGGAGAGCCGAAAGAACTGCCCCCAAGCTGTAGCATCAAAACAGAGAA 1380
Qy 3393 tgaatactcaacaaataggacaccaggaagaaaaagacatcttctctgaggagaatgt 3452
Db 1381 TGAATACTCAACCAATAGACACACAGGAAAAAAGACATCTTCTCTCAGGAGATGT 1440
Qy 3453 gcgtgctcctataactcaagttaataactccagcaaccttttaacatcacagcagaggt 3512
Db 1441 GCCTGGCTCTATAAATCAAGTAAATTAATCTTCCAGCAACCTTTAAACATCAGCAGAGGT 1500
Qy 3513 ttatcctggaggtttagaccacccagctcaaccaatgctggaagaagttagctttacc 3572
Db 1501 TTGTCTCTGGAGTTTGAGACCCCGAGCTCAACCAAAATGCTGGAGAAAGTGTAGCTTTACC 1560
Qy 3573 tgctctctctgcttaagtccaatagatagcagggcccttagaagaagaagatctggga 3632
Db 1561 TGCTCTCTCTAGCTAGTCAATAGATAGCAGGCGCTTAGGAAAGAGAGGTCTCGGA 1620
Qy 3633 tagtttaagtgtag 3648
Db 1621 TAGTTTAAAGTGAG 1636

RESULT 2

AL355587
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AL355587 151218 bp DNA HTG 15-APR-2001
Homo sapiens chromosome 10 clone RP11-561H23, *** SEQUENCING IN
PROGRESS ***, 9 unordered pieces.
AL355587
AL355587.8 GI:13660940
HTG; HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Burton, J.
Direct Submission
Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:13624988.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA561H23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 147891 bases at least Q40
Consensus quality: 149178 bases at least Q30
Consensus quality: 149893 bases at least Q20
Insert size: 150418; sum-of-contigs
Quality coverage: 4.69x in Q20 bases; sum-of-contigs Quality
coverage: 5.08x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 21608: contig of 21608 bp in length
* 21609 21708: gap of 100 bp
* 21709 48919: contig of 27211 bp in length
* 48920 49019: gap of 100 bp
* 49020 57539: contig of 8520 bp in length
* 57540 57639: gap of 100 bp
* 57640 69961: contig of 12322 bp in length
* 69962 70061: gap of 100 bp
* 70062 101299: contig of 31238 bp in length
* 101300 101399: gap of 100 bp
* 101400 114233: contig of 12834 bp in length
* 114234 114333: gap of 100 bp
* 114334 126584: contig of 12251 bp in length
* 126585 126684: gap of 100 bp
* 126685 138216: contig of 11532 bp in length
* 138217 138316: gap of 100 bp
* 138317 151218: contig of 12902 bp in length.

FEATURES
Source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-561H23"
/clone_lib="RP11-11.2"
1. .21608
/note="assembly_fragment:00445
fragment_chain:1
clone_end:7

misc_feature

```
misc_feature      vector_side:left"
21709..48919
/note="assembly_fragment:00191
fragment_chain:1"
49020..57539
/note="assembly_fragment:01755
fragment_chain:1"
57640..69961
/note="assembly_fragment:00689
fragment_chain:1"
70062..101299
/note="assembly_fragment:01153
fragment_chain:2"
101400..114233
/note="assembly_fragment:01618
fragment_chain:2"
114334..126584
/note="assembly_fragment:02159
fragment_chain:2"
126685..138216
/note="assembly_fragment:02086
fragment_chain:2"
138317..151218
/note="assembly_fragment:00492
fragment_chain:2
clone_end:SP6
vector_side:right"
BASE COUNT 45266 a 28197 c 29184 g 4771 t 800 others
ORIGIN

Query Match      41.2%; Score 1504.4; DB 79; Length 151218;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1505; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2143 cgggacgaactgaaacactctatcccaactggaatatataaaagaagaagatgac 2202
Db 89801 CAGGACGAGCTGAANAACCTCTATGCCAACCTGGAAATATATAAAGAAAGAGATGATC 89860
Qy 2203 acaaaacacccacactccagaaaaagcgtgctcgaagaaggcgtagctgttccatc 2262
Db 89861 ACAACAACCCCACTCCAGAAAAGCGTGCTCGAAGAAGGCGCTAGGTGCTTCCATC 89920
Qy 2263 atgagacgattacggagatcccgagacagtcagccggcagtcgtctaaagagacaa 2322
Db 89921 ATCAGACGCAATTACGGAGATCCAGAGACAGTCAGCCGCGAGTGTCTAAAGAGACAAG 89980
Qy 2323 gaggcgccgacatgacacagcgaaggcactgcccctcatcaggaagaacccccagag 2382
Db 89981 GAGGCGCGCGACCTGGCAACGAAGGCACTGCCCTCATCAGGAAGAACCCCGCAGAG 90040
Qy 2383 tcttcagggaacacaggggaaatcccaaggaggagaccctgaaacccgagctcttccatc 2442
Db 90041 TCTTCAGGGAACACAGGGAATCCAGAGGAGACCCCTGAAAACCGAGCTTCTCACTC 90100
Qy 2443 aagaatccacagacattatgaccacgtgagagaccacaaaggaagtcbaagtccta 2502
Db 90101 AAGAAATCCACAGACACTTATCACCACGCTGAGAGACCAACAGGAAGTCCAGTAGCCTA 90160
Qy 2503 ccacagaaagccagagagagacacagaaaaattccacatgaatccctgtcgagt 2562
Db 90161 CCCACGAANAGCCAGAGGAGAGACACAGAAAATTCACACTGGAATCCCTGCGGGT 90220
Qy 2563 aaaaaactaacacaaaaataaagaagacagcgggctgagtcacacgaggtcggtgcg 2622
Db 90221 AAAAACTAACACAAAAAATAAAGAAAGACAGCAGGAGCTGAGTCCACGAGTCGGTCCG 90280
Qy 2623 ttggtgtcgaatcagcagcgtcacacactcagtcagagaagaacactg99caccca 2682
Db 90281 TTGGGTGTCAAGTCAGCAGCGCTCACACCTCAGCTCAGAGAAAGAACTGGGCACCCA 90340
Qy 2683 cgaacatcgatgtacagaagtctcagtcagtcagcaagcgaaggaagactctt 2742
|||||
```

RESULT 3
AL355542/c
LOCUS
DEFINITION Homo sapiens chromosome 10 clone RP11-257J14, *** SEQUENCING IN
PROGRESS ***, 19 unordered pieces.
ACCESSION AL355542

AL355542 185146 bp DNA HTG 13-APR-2001

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

AL355542.9 GI:13624986
HTG; HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 185146)
Plumb,B.
Direct Submission
Submitted (12-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Apr 14, 2001 this sequence version replaced gi:13624506.

COMMENT

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA257J14
----- Summary Statistics
Assembly program: XGAP4; Version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 175310 bases at least Q40
Consensus quality: 178563 bases at least Q30
Consensus quality: 180883 bases at least Q20
Insert size: 183346; sum-of-contigs
Insert size: 145596; 18.0% error; agarose-fp
Quality coverage: 4.36x in Q20 bases; sum-of-contigs Quality
coverage: 7.64x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 11771: contig of 11771 bp in length
* 11772 11871: gap of 100 bp
* 11872 61631: contig of 49760 bp in length
* 61632 61731: gap of 100 bp
* 61732 79327: contig of 17596 bp in length
* 79328 79427: gap of 100 bp
* 79428 83309: contig of 3882 bp in length
* 83310 83409: gap of 100 bp
* 83410 109528: contig of 26119 bp in length
* 109529 109628: gap of 100 bp
* 109629 112808: contig of 3180 bp in length
* 112809 112908: gap of 100 bp
* 112909 118905: contig of 5997 bp in length
* 118906 119005: gap of 100 bp
* 119006 133743: contig of 14738 bp in length
* 133744 133843: gap of 100 bp
* 133844 143213: contig of 9370 bp in length
* 143214 143313: gap of 100 bp
* 143314 158185: contig of 14872 bp in length
* 158186 158285: gap of 100 bp
* 158286 162061: contig of 3776 bp in length
* 162062 162161: gap of 100 bp
* 162162 164747: contig of 2586 bp in length
* 164748 164847: gap of 100 bp
* 164848 167658: contig of 2811 bp in length
* 167659 167758: gap of 100 bp
* 167759 170009: contig of 2251 bp in length
* 170010 170109: gap of 100 bp
* 170110 173787: contig of 3678 bp in length
* 173788 173887: gap of 100 bp
* 173888 176350: contig of 2463 bp in length
* 176351 176450: gap of 100 bp
* 176451 179441: contig of 2991 bp in length

* 179442 179541: gap of 100 bp
* 179542 182272: contig of 2731 bp in length
* 182273 182372: gap of 100 bp
* 182373 185146: contig of 2774 bp in length.

FEATURES

source

1..185146 Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone_lib="RP11-257J14"
/clone_lib="RP11-11.1"
1..11771
/note="assembly_fragment:00155
fragment_chain:1"
11872..61631
/note="assembly_fragment:00988
fragment_chain:1"
61732..79327
/note="assembly_fragment:01331
fragment_chain:1"
79428..83309
/note="assembly_fragment:03102
fragment_chain:1"
83410..109528
/note="assembly_fragment:02919
fragment_chain:1"
109629..112808
/note="assembly_fragment:01757
fragment_chain:1"
112909..118905
/note="assembly_fragment:01162
fragment_chain:1"
119006..133743
/note="assembly_fragment:01246
fragment_chain:1"
133844..143213
/note="assembly_fragment:01266
fragment_chain:1"
143314..158185
/note="assembly_fragment:03087
fragment_chain:1"
158286..162061
/note="assembly_fragment:00095"
162162..164747
/note="assembly_fragment:00400"
164848..167658
/note="assembly_fragment:00509"
167759..170009
/note="assembly_fragment:01477"
170110..173787
/note="assembly_fragment:01903"
173888..176350
/note="assembly_fragment:02102"
176451..179441
/note="assembly_fragment:02396"
179542..182272
/note="assembly_fragment:02514"
182373..185146
/note="assembly_fragment:03322"

BASE COUNT 50876 a 38769 c 39206 g 54480 t 1815 others
ORIGIN

Query Match 39.6%; Score 1443; DB 79; Length 185146;
Query Local Similarity 99.1%; Pred. No. 0;
Matches 1493; Conservative 0; Mismatches 10; Indels 4; Gaps 4;

Qy 2143 cgggacgagctgaaaaactctatgcccaactggaatataataaagaagaatgac 2202
Db 164377 CAGGACGAGCTGAAAAAATCTATGCCCACTGGAATATATAAAGAAACAGATGATC 164418
Qy 2203 acaaacacccccaccctccagaaaaaacggtgctcgaagaaggcctaggtcggtccatc 2262
|||||


```
/clone="Qcce-20671"
/sex="male"
/tissue_type="cerebellum cortex"
/clone_lib="macaque brain cDNA library Qcce"
/dev_stage="adult"
213..1166
CDS
/codon_start=1
/product="hypothetical protein"
/protein_id="BAB19001.1"
/db_xref="GI:11611573"
/translation="MLQKSLSVIASAKERTGLAGTGTAGVEESAKSOKPLPKDRET
NRHNSDNTETKDPFONSNAPEPRKPKSGIMKQORVNPPTANSDLNPGATOMKD
NFDIGEVCEVYDLTPGPVPSKVKHVSIVASEMENPTLSLKEKSHRKPAAEV
CQSNQKCTDKASVENCWLGIOGQIILEDKHFISKTPVLERKEENGQOPHAAKVCAG
QSEELPPKAVASVENLNQIGHKEKTSFEENVRSNNFNFOPLMSRAEVCWP
EFTPAQPNAGRVALPVSALSASIRAPRKEEVDTEKV"
BASE COUNT      629 a 395 c 414 g 419 t
ORIGIN

Query Match      29.5%; Score 1077.8; DB:85; Length 1837;
Best Local Similarity 95.7%; Pred. No. 7.6e-242;
Matches 1119; Conservative 0; Mismatches 47; Indels 3; Gaps 1;

QY 2480 aaacggaagtcacgtacgtacccacagaaagccaaagaggagagacacagaaatt 2539
Db 1 AAACGGGAAGTCCAGTAGGCTACCCACAGAAAGCCAAAGAGAGAGGTGACAGAAAATT 60
QY 2540 ccacactggaatccctgtcgggttaaaactaacacaaaactaaagaagacagcgagg 2599
Db 61 CCACACTGGAATCCTGTCGAGTAAGAATACTGACACAAAATACTAAAGAGACACTGAGG 120
QY 2600 ctgagtcacggagtcggtgctgtgtgtaagtcagcagcgctcaaacctcagct 2659
Db 121 CTGAGTCCACGGAGTCGCTGCGTGTGTGTCGAAGTCAGCAAGTGCTCACAACTCAGCT 180
QY 2660 cagagaagaataactggccacacagacatgatgttacagaagctctcagtgctatag 2719
Db 181 CAGAGAAGAAGCCCTGGACACCCACGACGTCATGTTTACAGAAGTCTCTCAGTGTCATAG 240
QY 2720 caagcgcaaggaagactcttgattagctgggaaaccccaacacagcaggtgtgaag 2779
Db 241 CAAGTGGCCAGGAGAACTCTTGGATTAGCTGGGAAACCCCAACACAGCGCTGGGAAG 300
QY 2780 aacgcactaatcccgagaacacttggccaaagataaagagacaaacagaaatcacctcaa 2839
Db 301 AAAGTGCTAAATCCAGAAACCTCTGCCAAAGATAAAGAGACAAACAGAAATCACTCAA 360
QY 2840 attctgatacacagagactaaagatcctgcccccaaaactcaaatcctcgaggagc 2899
Db 361 ATTCTGATAACAGAGACTAAAGATCCTACCCCCCAAACTCAAAATCCTCGGAGGAGC 420
QY 2900 caagaaagcctcagaaatctgggattatgaacacaaaggggtcaacccaccactgcca 2959
Db 421 CAAGAAAGCCTCAGAAATCTGGGATTATGAACAAACAAAGGGTCAACCCCACTGCCA 480
QY 2960 attctgacctgaacccaggccaccacccagatgaaggacaaactttgacattggggaggtg 3019
Db 481 ATTCTGACCTGAACCCAGGCGCCACCAGATGAAGGACAACTTTGACATTGGGGAGGTGT 540
QY 3020 gtccctgggaggttatgacctgacctgctgtgtgcttccttcagaatacaaaagttcaaa 3079
Db 541 GTCCCTGGGAGGTTTATGACCTGACCCCTGTCTGCTGCTTCAAGATCAAAAGTTCAAA 600
QY 3080 agcagctatctattgtgctctgaaatggagaaaaacccacttttctcttaagagaga 3139
Db 601 AGCATGATATCTATTGTGGCTTCTGAATGGAGAAAAACCCCACTCTTTCTCTTAAGGAGA 660
QY 3140 aatctcaccacaagccttaaggcagctgaggtttgtcagcaatccaatcagaagcgcatag 3199
Db 661 AATCTCACCACAGCCCTAAGGCAGCTGAAGTTTGTGAGCAATCCCAATCAGAAGTGCACAG 720
QY 3200 ataaggctgaagtatgcctttgggagagccaaagcccgctccattttgggaagatgagaagc 3259
```

```
Db 721 ATAAGCGTGAAGTATGCTTTGGGGGATCCAAAGCCCAAGTCCATTTTGAAGATGAGAAGC 780
QY 3260 ttttgattccaagactccagtlctccagagagggcacaagaggaacacgaggtcagc 3319
Db 781 ATTTCATTTTCAAGACTCCAGTTCTCCAAAGAGAGGGCAAAAGAGAGAGCAACGAGGTGAGC 840
QY 3320 ctgctgcagccaatgtgtgtcgtgggcagagcgcaagaactgcccccaaaagctgtagcat 3379
Db 841 CTCATGCAGCCAAAGTGTGTGCTGGGCAGAGTGAAGAACTGCCCCCCTCAAGCTGTAGCAT 900
QY 3380 caaaaacagagaatgaaatctcaaccataatagacacccagggaaaaaaagacatctctt 3439
Db 901 CAAAAACAGAGAAATGAAAAATCTCAACCAATATAGACACACCAGGAAAAAAGACA---TCCT 957
QY 3440 ctgagggagaatgtcgtggtgctcctataactcaagtcaataactctcagcaacctttaacat 3499
Db 958 TTGAGGAGAAATGCGTGGCTCTCTATAACTCAAGTATAACTTCCAGCAACCTTTAATGT 1017
QY 3500 caagagcagaggtttgtccttgggagtttgagaccccgctcaaccaaactgtcgaagaa 3559
Db 1018 CACGAGCAGAGGTGTGCTTGGAGTTTGAGACCCCGAGCTCAACCAAAATGCTGGAAGAA 1077
QY 3560 gtgtagctttacctgcctcttctgtccttaagtgcacaaataagatagcagggccttaggaag 3619
Db 1078 GTGTAGCTTTTACCTGTTTCTTCTCTCTAAAGTCAAGTAAAGATAGCAGGGCCTCGGAAG 1137
QY 3620 aagagatctgggagtagttttaaagtgtag 3648
Db 1138 AAGAAGTCTGGGATACCTTTTAAAGTGTAG 1166
```

```
RESULT 5
AL358932
LOCUS      AL358932      111117 bp      DNA      HTG      23-JAN-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-395P8, *** SEQUENCING IN
PROGRESS ***, 25 unordered pieces.
ACCESSION  AL358932
VERSION    AL358932.3
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Sims,S.
REFERENCE  1 (bases 1 to 111117)
            Direct Submission
            Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
            On Aug 27, 2000 this sequence version replaced gi:9926667.
            ----- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----- Project Information
            Center project name: BA395P8
            ----- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 90261 bases at least Q40
            Consensus quality: 99226 bases at least Q30
            Consensus quality: 104490 bases at least Q20
            Insert size: 108717; sum-of-contigs
            Insert size: 178772; 24.0% error; agarose-fp
            Quality coverage: 1.84x in Q20 bases; sum-of-contigs Quality
            coverage: 1.42x in Q20 bases; agarose-fp
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
```



```
QY 301 ggcgcgtacagttggcgggctgcgcgggaaagtggccagcccttgccagcgccacccc 360
|||||
Db 83703 ggcgcgtacagttggcgggctgcgcgggaaagtggccagcccttgccagcgccacccc 83762
|||||
QY 361 tcctgtaccggtggtggacacactgacacgacgaccccaacttctcaacgtgactgtg 420
|||||
Db 83763 TCCTTGGACCGGCGTGGACACACTGACACAGCCACCAACTTCCTCAACAGTGTGCTG 83822
|||||
QY 421 cagagcaataagtgcggggagcagaacttcgagagacaccttgattggtaccaggcgctg 480
|||||
Db 83823 CAGAGCAATAAGTCGCGGGAGCAGAACTTCAGGACGACCTGATGGTACCAGGCGCTG 83882
|||||
QY 481 gtgtgagccttctgaggcgagccagcatctcccgggcgccatcaacttcagcacc 540
|||||
Db 83883 GTGTGGAGCCTTCTGGAGGCGGAGCCAGCATCTCCGGGCGGCCATCACCCTTCAGCACC 83942
|||||
QY 541 gattcgctgccacacggccccacagttcttctccagcccgacgacgagagagccgc 600
|||||
Db 83943 GATTCGCTGTCCGACCGCGCCACAGGCTCTTCTCCAGGCCACGCGGAGGAGCGCG 84002
|||||
QY 601 atcgtctcaagaacctgtctctccgaccccaactgcccacacgcaactcttgagacc 660
|||||
Db 84003 ATCTGTCTCAAGACCTGTCTCTCTCGCACCCACCTGCGCAAGGCACTCTGGAGACC 84062
|||||
QY 561 gattgttccacggctcccgcaagtggagggcccccacttaccacgcccggccccaat 720
|||||
Db 84063 GAGTGTGTTCACGCGCTTCGCGGCGCAAGTGGAGGCGCCCACTTACACCGCGCGCCCAAT 84122
|||||
QY 721 caggggcccgggcctggccacagctggcgccgaagacagggctcgaggggacaag 780
|||||
Db 84123 CAGGGGCGCGGCGCTTGGCGCCACAGCTGGCGGCGCAAGGCGCTCGGCGGGGCAAG 84182
|||||
QY 781 agccactcaagtgtctccgcttattctgagtgagagacgggagttacaagcccggg 840
|||||
Db 84183 AGCCACTCAAGTGGTCTCGCCTTATCTGAGTGCAGAACGGGAGTTACAAGCCCGGG 84242
|||||
QY 841 tggctgtacttctcttgccatctacgggttgagcctcaacttggtcccggaattc 900
|||||
Db 84243 TGGCTGTACTTCTTCTCTGCCATCTAGCGGTTGAGCGCTAACCTGTGTCGCGGATTC 84302
|||||
QY 901 agg 903
|||
Db 84303 AGG 84305

RESULT 6
AL139821 199526 bp DNA HTG 17-MAR-2001
Homo sapiens chromosome 10 clone RP11-80K21, *** SEQUENCING IN
PROGRESS ***, 6 unordered pieces.
ACCESSION AL139821
VERSION AL139821
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Plumb, B.
Direct Submission
Submitted (15-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 20, 2001 this sequence version replaced gi:10185474.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA80K21
----- Summary Statistics
```

Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 197300 bases at least Q40
Consensus quality: 198299 bases at least Q30
Consensus quality: 198676 bases at least Q20
Insert size: 199026; sum-of-contigs
Quality coverage: 7.42x in Q20 bases; sum-of-contigs Quality
coverage: 10.32x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 37012: contig of 37012 bp in length
* 37013 37112: gap of 100 bp
* 37113 163560: contig of 126448 bp in length
* 163561 163660: gap of 100 bp
* 163661 171563: contig of 7903 bp in length
* 171564 171663: gap of 100 bp
* 171664 186786: contig of 15123 bp in length
* 186787 186886: gap of 100 bp
* 186887 196057: contig of 9171 bp in length
* 196058 196157: gap of 100 bp
* 196158 199526: contig of 3369 bp in length.

FEATURES

source
1..199526

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-80K21"
/clone_lib="RPC1-11.1"

misc_feature
1..37012
/note="assembly_fragment:00668"
fragment_chain:1
clone_end:T7

misc_feature
37113..163560
/note="assembly_fragment:02214"
fragment_chain:1
clone_end:SP6

misc_feature
163661..171563
/note="assembly_fragment:00510"
171664..186786
/note="assembly_fragment:01538"

misc_feature
186887..196057
/note="assembly_fragment:01931"
196158..199526
/note="assembly_fragment:03255"

BASE COUNT 57428 a 38898 c 39349 g 63347 t 504 others
ORIGIN

Query Match 24.8%; Score 903; DB 79; Length 199526;
Best Local Similarity 100.0%; Pred. No. 4.6e-201;
Matches 903; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggagagccatggcttacccttactctctgctctgcttctgctcagctgggattggga 60
|||||

Db 97153 ATGGAGACCATGGCTTACCCCTTACTCTCTGCTCTGCTGCTGCTGCTGCTGCTGGA 97217
|||||

QY 61 gctgtgtggccagcgccgaccccaaggagcgccgattccctcgagagagaccccg 120
|||||

Db 97218 GCTGTTGGCGCAGCGCGACCCCAAGAGAGCGCGGATTCCTCTCGAGAGAGACCCCG 97277
|||||

QY 121 aaggggaagcgcaagcccgagcgccgggtcgagcctctgctcgagctcgctcgctcc 180
|||||

Db 97278 AAGGGGAAGCCGACCCAGCAGCGGGTGGAGCCCTCTGCTCGACATCTCTGGCTGCC 97337
QY 181 tggagcgcgtccaccgatggcaccatctggcgagaaactgcgcgagaggtgcccattg 240
Db 97338 TGGAGCGGCTCCACCGATGGCACCATTCTGGCGCAGAACTGCCGAGAGGTGCCCATG 97397
QY 241 gacgtggcctcttacctctacacccgggagctccaccacagctgaagcgagcccaactgctcc 300
Db 97398 GAGTGGCCTCTTACCTCTACACGGGGACTCCACACAGCTGAAGGAGCCAACTGCTCC 97457
QY 301 gcccgtacagttggcgggcctgcggggaagtggcagccctggcagcggaacccc 360
Db 97458 GGGCGGTACAGTTGGCGGGCTTGGCGGGAAGTGGCCAGCCCTGGCCAGCGCACCC 97517
QY 361 tccctgacccggcgctggacacactgacacgcgcacacacacttctcaactgatgctg 420
Db 97518 TCCTTGCACCGGGCGGTGACACACTGACACAGCCACCACTTCTCAACTGATGCTG 97577
QY 421 cagagcaataagtgcgggagcagaacttgcaggagcactgattggtaccagcgctg 480
Db 97578 CAGAGCAATAAGTCCGGGGAGCAGAACTTGCAGGAGCAGCTGGATTGGTACCAGGCGCTG 97637
QY 481 gigtgagcctcttgaggcgagccagcatctcccgggcgccatcaacttcagcacc 540
Db 97638 GTCGTGAGCCTTCTGGAGGGCGAGCCAGCATCTCCGGGGGCCCATCACTTCAGCAACC 97697
QY 541 gattcgtgctgcgcacggcccccacaggtcttctctccagccacgcgcgagagagcgccg 600
Db 97698 GATTCGCTGTCGCACCGGCCCCACAGGCTTCTCTCCAGGCCACGCGGAGAGCGCGC 97757
QY 601 atctgctccaagactgtctctctccgcacccacccacttggccaaagcactctggagacc 660
Db 97758 ATCTGCTCCAAGACTGTCTCTCTCCGACCCACCTGGCCAGCCACTCTGGAGACC 97817
QY 661 gactggtccagcgcctccgagcgaagtggagggccacacttaccacgcgcgcccccaat 720
Db 97818 GAGTGGTTCACCGGCTCCGGCGCAAGTGGAGGGCCCACTTACCGCGCGGCCCAAT 97877
QY 721 caggggcccgccggcctggcgccacagctggcgcgcaagagcgctcgcggggacaaag 780
Db 97878 CAGGGCGCCGGCGCTGGGCCACAGCTGGCGGCGCAAGGAGCGGCTCGCGGGGACAG 97937
QY 781 agccacttaagtgtctcgccttatcttggtgcgagaaagcgagttacaaagccccggg 840
Db 97938 AGCCACTTCAAGTGTCTCGCCTTATCTGAGTGGAGAGCGGAGGTACAAAGCCCGG 97997
QY 841 tgcgtggtactcttctctccatctacggttgcagcctaacctgctccggaattc 900
Db 97998 TGGCTGGTTACTTCTCTCTCCATCTACGGGTTGCAGCCTTAACCTGGTCCCGGAATTC 98057
901 agg 903
Db 98058 AGG 98060
RESULT 7
AL359974/c
LOCUS
DEFINITION Homo sapiens chromosome 10 clone RP13-236A4, *** SEQUENCING IN
PROGRESS ***, 25 unordered pieces.
ACCESSION AL359974
VERSION AL359974.4 GI:9864577
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
i (bases 1 to 143878)
Sims,S.
Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk

COMMENT

On Aug 22, 2000 this sequence version replaced gi:9795088.
----- Genome Centre
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba236A4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 130224 bases at least Q40
Consensus quality: 135875 bases at least Q30
Consensus quality: 138890 bases at least Q20
Insert size: 141478; sum-of-contigs
Insert size: 205930; 23.5% error; agarose-fp
Quality coverage: 3.17x in Q20 bases; sum-of-contigs Quality
coverage: 2.50x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 12130: contig of 12130 bp in length
* 12131 12230: gap of 100 bp
* 12231 12274: contig of 8044 bp in length
* 20275 20374: gap of 100 bp
* 20375 23062: contig of 2688 bp in length
* 23063 23162: gap of 100 bp
* 23163 26227: contig of 3065 bp in length
* 26228 26327: gap of 100 bp
* 26328 33260: contig of 6933 bp in length
* 33261 33360: gap of 100 bp
* 33361 35571: contig of 2211 bp in length
* 35572 35671: gap of 100 bp
* 35672 38834: contig of 3163 bp in length
* 38835 38934: gap of 100 bp
* 38935 44371: contig of 5437 bp in length
* 44372 44471: gap of 100 bp
* 44472 47729: contig of 3258 bp in length
* 47730 47829: gap of 100 bp
* 47830 52657: contig of 4828 bp in length
* 52658 52757: gap of 100 bp
* 52758 54776: contig of 2019 bp in length
* 54777 54876: gap of 100 bp
* 54877 57327: contig of 2451 bp in length
* 57328 57427: gap of 100 bp
* 57428 60599: contig of 3172 bp in length
* 60600 60699: gap of 100 bp
* 60700 64861: contig of 4162 bp in length
* 64862 64961: gap of 100 bp
* 64962 71930: contig of 6969 bp in length
* 71931 72030: gap of 100 bp
* 72031 75423: contig of 3399 bp in length
* 75430 75529: gap of 100 bp
* 75530 78541: contig of 3012 bp in length
* 78542 78641: gap of 100 bp
* 78642 80961: contig of 2320 bp in length
* 80962 81061: gap of 100 bp
* 81062 83093: contig of 2038 bp in length
* 83100 83199: gap of 100 bp
* 83200 95484: contig of 12285 bp in length
* 95485 95584: gap of 100 bp
* 95585 97735: contig of 2151 bp in length
* 97736 97835: gap of 100 bp
* 97836 102251: contig of 4416 bp in length
* 102252 102351: gap of 100 bp
* 102352 131392: contig of 29041 bp in length

```
* 131393 131492: gap of 100 bp
* 131493 137581: contig of 6089 bp in length
* 137582 137681: gap of 100 bp
* 137682 143878: contig of 6197 bp in length.
FEATURES
    source
        1. .143878
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="10"
            /clone="RP13-236A4"
            /clone_lib="RPC1-13.1"
        1. .12130
            /note="assembly_fragment:01047"
            fragment_chain:1
            12231..20274
            /note="assembly_fragment:00007"
            fragment_chain:1
            20375..23062
            /note="assembly_fragment:01237"
            fragment_chain:1
            23163..26227
            /note="assembly_fragment:01103"
            fragment_chain:1
            26328..33260
            /note="assembly_fragment:00266"
            fragment_chain:1
            33361..33571
            /note="assembly_fragment:01006"
            fragment_chain:1
            35672..38834
            /note="assembly_fragment:00201"
            fragment_chain:2
            38935..44371
            /note="assembly_fragment:01057"
            fragment_chain:2
            44472..47729
            /note="assembly_fragment:00179"
            fragment_chain:2
            47830..52657
            /note="assembly_fragment:00444"
            fragment_chain:3
            52758..54776
            /note="assembly_fragment:00414"
            fragment_chain:3
            54877..57327
            /note="assembly_fragment:00586"
            fragment_chain:4
            57428..60599
            /note="assembly_fragment:00641"
            fragment_chain:4
            60700..64861
            /note="assembly_fragment:00717"
            fragment_chain:5
            64962..71930
            /note="assembly_fragment:01116"
            fragment_chain:5
            72031..75429
            /note="assembly_fragment:00156"
            75530..78541
            /note="assembly_fragment:00233"
            78642..80961
            /note="assembly_fragment:00273"
            81062..83099
            /note="assembly_fragment:00279"
            83200..95484
            /note="assembly_fragment:00515"
            95585..97735
            /note="assembly_fragment:00654"
            97836..102251
            /note="assembly_fragment:00675"
            102352..131392
            /note="assembly_fragment:00719"
            131493..137581
```

```
misc_feature      /note="assembly_fragment:00767"
137682..143878
/note="assembly_fragment:00852
clone_end:SP6
vector_side:right"
BASE COUNT      44271 a 25988 c 25788 g 45414 t 2417 others
ORIGIN

Query Match      6.2%; Score 227.2; DB 80; Length 143878;
Best Local Similarity 93.0%; Pred. No. 1.2e-42;
Matches 238; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 1112 gaaggggtccggtacagcatatttcaggaagtcacaaagatgtgtcagaagaagcctatg 1171
|||||
Db 89285 GAAGGGGTCCGGATCAGCATATTTTCAGGAAGTACAAAGATGTGTCTAGAGAAGCCTATG 89226

Qy 1172 tctgcctaccttcagggagggtgcccttctgtctgatgacagccatgcttcgtcc 1231
|||||
Db 89225 TCTGCCTACCTTCAGGGAGGGCTGCCCTTCTGTGCTGATGACAGCCCATGCTTCGTCC 89166

Qy 1232 aggaagataagtatttcagcttgcctcatctcttcacagccttcgcaagcaagcctcggcctcg 1291
|||||
Db 89165 AGGAAGATAAGTATTTCAGCTTGCCTATCTCTCTCCAAAGCCTGTGTATGTGCTCG 89106

Qy 1292 acttcgttagcatgtgtgtctaccacttttcgcaagccttcgcaagcctcggcctcg 1351
|||||
Db 89105 ACTTCGTTAGCATGTGTGTGCTTACCACCTTTCGCAAGCAAGGTAAACCCAGCAACCC 89046

Qy 1352 gccttacctgttggg 1367
|||||
Db 89045 TGGTATGATCCTGTA 89030

RESULT 8
LOCUS      AL161654 159596 bp DNA HTG 11-APR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-59G22, *** SEQUENCING IN
PROGRESS ***, 8 unordered pieces.
ACCESSION      AL161654
VERSION      AL161654.8 GI:13620309
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 159596)
AUTHORS      Burton,J
TITLE      Direct Submission
JOURNAL
COMMENT
On Apr 12, 2001 this sequence version replaced gi:13567947.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA59G22
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 157954 bases at least Q40
Consensus quality: 158428 bases at least Q30
Consensus quality: 158729 bases at least Q20
Insert size: 158896; sum-of-contigs
Insert size: 159333; 6.0% error; agarose-fp
Quality coverage: 6.69x in Q20 bases; sum-of-contigs Quality
coverage: 6.67x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
```

```
/clone_lib="RpCI-11.1"
```

100

```
misc_feature 1..149917
/note="assembly_fragment:00886.0"
misc_feature 150018..153565
/note="assembly_fragment:01186"
misc_feature 153666..162836
/note="assembly_fragment:01998"
misc_feature 162937..186964
/note="assembly_fragment:00147
vector_side:right"
vector_end:SP6

BASE COUNT 64194 a 34055 c 33553 g 54858 t 304 others
ORIGIN

Query Match 3.1%; Score 112; DB 79; Length 186964;
Best Local Similarity 95.8%; Pred. No. 1.1e-15;
Matches 115; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1002 ttcagagtgtatgccaattaaagccctaggatttcgtttcttgagccctatgagtcatttg 1061
|||||
179738 TTCTAGTGTATGCCAATTAAAGGCCCTAGGATTCGTTCTTGAGGCCCTATGAGTCATTTG 179797
|||||

Qy 1062 caaagcaggattctatccctcgttgagcttaccagtgaaacaacttcggagaagggtcc 1121
|||||
Db 179798 CAAGCAGGATTCATCATCTCTGGAGCTTACCAGTCAACAACCTTTCGGAGTAAGTCCCC 179857
|||||

RESULT 10
AL445244/c
LOCUS AL445244 160390 bp DNA HTG 20-JAN-2001
DEFINITION Homo sapiens chromosome 10 clone RP13-99J21, *** SEQUENCING IN
PROGRESS ***, 6 unordered pieces.
ACCESSION AL445244
VERSION AL445244.4 GI:10800726
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sims,S.
Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Oct 13, 2000 this sequence version replaced gi:10798428.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bB99J21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 158004 bases at least Q40
Consensus quality: 159033 bases at least Q30
Consensus quality: 159555 bases at least Q20
Insert size: 159890; sum-of-contigs
Insert size: 144185; 13.1% error; agarose-fp
Quality coverage: 5.59x in Q20 bases; sum-of-contigs Quality
coverage: 6.20x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
* 1 43907: contig of 43907 bp in length
* 43908 44007: gap of 100 bp
* 44008 55206: contig of 11199 bp in length
* 55207 55306: gap of 100 bp
* 55307 67735: contig of 12429 bp in length
* 67736 67835: gap of 100 bp
* 67836 73636: contig of 5801 bp in length
* 73637 73736: gap of 100 bp
* 73737 100153: contig of 26417 bp in length
* 100154 100253: gap of 100 bp
* 100254 160390: contig of 60137 bp in length.
FEATURES
Location/Qualifiers
source 1..160390
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP13-99J21"
/clone_lib="RPCI-13.1"
misc_feature 1..43907
/note="assembly_fragment:01136
clone_end:T7
vector_side:left"
misc_feature 44008..55206
/note="assembly_fragment:01384
fragment_chain:1"
misc_feature 55307..67735
/note="assembly_fragment:01898
fragment_chain:1"
misc_feature 67836..73636
/note="assembly_fragment:01997
fragment_chain:1"
misc_feature 73737..100153
/note="assembly_fragment:01657
fragment_chain:1"
misc_feature 100254..160390
/note="assembly_fragment:01063
fragment_chain:1
clone_end:SP6
vector_side:right"
BASE COUNT 55825 a 29765 c 28364 g 45935 t 501 others
ORIGIN

Query Match 3.0%; Score 108; DB 81; Length 160390;
Best Local Similarity 100.0%; Pred. No. 1e-14;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 901 aggggtgtcatgaaagtgcataaattcttcagaaagtggaattgacccaattgctcaagt 960
|||||
Db 136464 AGGGGTGTCTGAAAGTTTGACATAAATCTTCAGAAAGTGGACATTGACCAATGCTCAAGT 136405
|||||

Qy 961 gatggtgtgttttcagggaactcataaattgccacctcaacaatttcagag 1008
|||||
Db 136405 GATGCTGTGTTTTTCAGGAACCTCATAAATGCCCTCAACAATTCAGAG 136357
|||||

RESULT 11
I66494/c
LOCUS I66494 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
Location/Qualifiers
source 1..7218
/organism="unknown"
```



```

repeat_unit
complement(1287..1346)
/notes="55 bp imperfect inverted repeat"
gene
1648..3156
/genes="SC5H1.40c"
CDS
1648..3156
/genes="SC5H1.40c"
/notes="SC5H1.40c, possible regulatory protein, len: 502
aa; similar to e.g. SW:SPI5_STRGR (EMBL:M32687)
Streptomyces griseus sporulation protein (529 aa), fasta
scores; opt: 394 z-score: 452.2 E(): 7.2e-18, 27.9%
identity in 530 aa. Similar to SC7A1.26 (EMBL:AL034447)
S.coelicolor possible transcriptional regulator (500 aa)
(43.0% identity in 491 aa overlap)"
/codon_start=1
/transl_table=11
/label="SC5H1.40c"
/product="putative regulatory protein"
/protein_id="CAB42965.1"
/db_xref="GI:4835343"
/db_xref="SPTREMBL:Q9X7U7"
/translation="MSKEPNTRLSDFLAGWSKGLARLVNRQAAAGHPOLATDTS
RYRWIDMGEIPRDPVPRVLAALFETRLGRVVTIEDLGVRRGRAGKRPHDGESEHPD
GWPADEPTAAVLFTETGMDLNLNREGLVAGAAALAGSALSAMIDWLTLPALAAD
ADLHOPADPADFDYEAAPIGSOEVEELERSVEVFRANDAAARGQKRAVYQOL
NEVGGIATYHPHQLRRLWGAANLAVLAGWMSHDVGLPTAQKTFVIAHAAREGG
DRPAGELASRAARQWHLGKPDALDLMLKLAQSGSGEQLPRTKAMLYTIEAWAQAS
MKGQAMRTTLAGRAEDLFYSDKADVPDPDMQOTFKREDYGMQALAYRTLAEFPGAA
AAHAYADKALALRIDGRORSKFYLSNASACFTADDPQADRYARLARLALVSGNSNS
QRTWDLROMYRLTAEYAGYKIQELREIKALPKGKSPRGKPGGTVAFA"
complement(3176..4492)
/genes="SC5H1.39"
complement(3176..4492)
/genes="SC5H1.39"
/notes="SC5H1.39, hypothetical protein, len: 438 aa;
unknown function, probable CDS suggested by positional
base preference, GC frame analysis and amino acid
composition"
/codon_start=1
/transl_table=11
/label="SC5H1.39"
/product="hypothetical protein"
/protein_id="CAB42964.1"
/db_xref="GI:4835342"
/db_xref="SPTREMBL:Q9X7U6"
/translation="MSSHLSADHPAAQSPQGSVEALISQTRRLRGDVAVRDSDDD
TTDPRGWRALYDLALHQLTDLDAHLAOLRDGPAPAPRAGSLLSRVGSAEWNLLTDE
AEWSAEFLHGRDPAVPLSDLEPLVLTEDRPKLTAMVTCLVDGPRIDGFEFV
RDPGVSRTVMHGEPLVDAGSTASWAVLRVDVGLRRSQRAVSETRSLRRPASPW
HGLRPPDGRPLDLAGAHLPTTSATGRSREGGWYDACPLGADGETLITVGDFFGLP
EADAYPAVLIGALRGALAGTRPGRRLRLAELRLATLAPVDGAVYCYRPTDTTL
TWAHASCAPALLFRGGTGHALSGPDGPGAGDGPGEATLEAGDLLLHTGALPEA
VEHLTLAPRLTAADSAGSDGVRVLTRELDFGFRQDACVLIARVTP"
complement(4765..5793)
/genes="SC5H1.38"
complement(4765..5793)
/genes="SC5H1.38"
/notes="SC5H1.38, probable myo-inositol dehydrogenase, len:
342 aa; similar to many e.g. SW:M12D_BACSU (EMBL:M76431),
idh, Bacillus subtilis myo-inositol 2-dehydrogenase (344
aa), fasta scores; opt: 864 z-score: 958.5 E(): 0, 38.1%
identity in 336 aa overlap. Also similar to SW:STRI_STRGR
(EMBL:Y00459), stri, Streptomyces griseus streptomycin
biosynthesis myo-inositol oxidoreductase (348 aa) (32.7%
identity in 343 aa overlap). Weakly similar to TR:O69945
(EMBL:AL023962) S.coelicolor possible oxidoreductase (430
aa) (29.8% identity in 228 aa overlap). Contains Pfam
match to entry PF01408 GFO_IDH_MocA, Oxidoreductase
family, score 153.50, E-value 3.6e-42"
/codon_start=1
/transl_table=11
/label="SC5H1.38"
/product="putative myo-inositol dehydrogenase"

regulation"
complement(1287..1346)
/notes="55 bp imperfect inverted repeat"
gene
1648..3156
/genes="SC5H1.40c"
CDS
1648..3156
/genes="SC5H1.40c"
/notes="SC5H1.40c, possible regulatory protein, len: 502
aa; similar to e.g. SW:SPI5_STRGR (EMBL:M32687)
Streptomyces griseus sporulation protein (529 aa), fasta
scores; opt: 394 z-score: 452.2 E(): 7.2e-18, 27.9%
identity in 530 aa. Similar to SC7A1.26 (EMBL:AL034447)
S.coelicolor possible transcriptional regulator (500 aa)
(43.0% identity in 491 aa overlap)"
/codon_start=1
/transl_table=11
/label="SC5H1.40c"
/product="putative regulatory protein"
/protein_id="CAB42965.1"
/db_xref="GI:4835343"
/db_xref="SPTREMBL:Q9X7U7"
/translation="MSKEPNTRLSDFLAGWSKGLARLVNRQAAAGHPOLATDTS
RYRWIDMGEIPRDPVPRVLAALFETRLGRVVTIEDLGVRRGRAGKRPHDGESEHPD
GWPADEPTAAVLFTETGMDLNLNREGLVAGAAALAGSALSAMIDWLTLPALAAD
ADLHOPADPADFDYEAAPIGSOEVEELERSVEVFRANDAAARGQKRAVYQOL
NEVGGIATYHPHQLRRLWGAANLAVLAGWMSHDVGLPTAQKTFVIAHAAREGG
DRPAGELASRAARQWHLGKPDALDLMLKLAQSGSGEQLPRTKAMLYTIEAWAQAS
MKGQAMRTTLAGRAEDLFYSDKADVPDPDMQOTFKREDYGMQALAYRTLAEFPGAA
AAHAYADKALALRIDGRORSKFYLSNASACFTADDPQADRYARLARLALVSGNSNS
QRTWDLROMYRLTAEYAGYKIQELREIKALPKGKSPRGKPGGTVAFA"
complement(3176..4492)
/genes="SC5H1.39"
complement(3176..4492)
/genes="SC5H1.39"
/notes="SC5H1.39, hypothetical protein, len: 438 aa;
unknown function, probable CDS suggested by positional
base preference, GC frame analysis and amino acid
composition"
/codon_start=1
/transl_table=11
/label="SC5H1.39"
/product="hypothetical protein"
/protein_id="CAB42964.1"
/db_xref="GI:4835342"
/db_xref="SPTREMBL:Q9X7U6"
/translation="MSSHLSADHPAAQSPQGSVEALISQTRRLRGDVAVRDSDDD
TTDPRGWRALYDLALHQLTDLDAHLAOLRDGPAPAPRAGSLLSRVGSAEWNLLTDE
AEWSAEFLHGRDPAVPLSDLEPLVLTEDRPKLTAMVTCLVDGPRIDGFEFV
RDPGVSRTVMHGEPLVDAGSTASWAVLRVDVGLRRSQRAVSETRSLRRPASPW
HGLRPPDGRPLDLAGAHLPTTSATGRSREGGWYDACPLGADGETLITVGDFFGLP
EADAYPAVLIGALRGALAGTRPGRRLRLAELRLATLAPVDGAVYCYRPTDTTL
TWAHASCAPALLFRGGTGHALSGPDGPGAGDGPGEATLEAGDLLLHTGALPEA
VEHLTLAPRLTAADSAGSDGVRVLTRELDFGFRQDACVLIARVTP"
complement(4765..5793)
/genes="SC5H1.38"
complement(4765..5793)
/genes="SC5H1.38"
/notes="SC5H1.38, probable myo-inositol dehydrogenase, len:
342 aa; similar to many e.g. SW:M12D_BACSU (EMBL:M76431),
idh, Bacillus subtilis myo-inositol 2-dehydrogenase (344
aa), fasta scores; opt: 864 z-score: 958.5 E(): 0, 38.1%
identity in 336 aa overlap. Also similar to SW:STRI_STRGR
(EMBL:Y00459), stri, Streptomyces griseus streptomycin
biosynthesis myo-inositol oxidoreductase (348 aa) (32.7%
identity in 343 aa overlap). Weakly similar to TR:O69945
(EMBL:AL023962) S.coelicolor possible oxidoreductase (430
aa) (29.8% identity in 228 aa overlap). Contains Pfam
match to entry PF01408 GFO_IDH_MocA, Oxidoreductase
family, score 153.50, E-value 3.6e-42"
/codon_start=1
/transl_table=11
/label="SC5H1.38"
/product="putative myo-inositol dehydrogenase"

misc_feature
complement(5062..5781)
/genes="SC5H1.38"
/notes="Pfam match to entry PF01408 GFO_IDH_MocA,
Oxidoreductase family, score 153.50, E-value 3.6e-42"
complement(5804..5808)
/notes="possible RBs"
complement(5917..7011)
/genes="SC5H1.37"
complement(5917..7011)
/genes="SC5H1.37"
/notes="SC5H1.37, hypothetical protein, len: 364 aa;
unknown function, similar to TR:P73120 (EMBL:D90903)
Synecocystis sp. hypothetical protein (358 aa), fasta
scores; opt: 492 z-score: 559.6 E(): 7.6e-24, 31.6%

Query Match 1.7%; Score 60.6; DB 3; Length 36583;
Best Local Similarity 50.2%; Pred. No. 0.0015;
Matches 150; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 220 ctgcgcagagagtgccatggaagtgccctcttacctctacacccggggagctccaccag 279
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31116 CTCGCCCTGGAGGACGCCATGGCCCTGCCACGCCCTGGGTGAGCGGACACCTCCCG 31175
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 280 ctgaagcgagcaactctccgcccgtacgagttgctgagcggcgtgagcggagagtgcca 339
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31176 GAGGCCCTCCGCCACTACGACGCGGACGCGGCGGCCCTCTGCTGCCGGTGCAGAGCGCG 31235
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 340 gccctggcgcgcgcacccctcttgcacggcgctgacacactgcacacgcgcacc 399
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31236 GCCCGGCTCAGCGCCAGTGGTACGAGAACCTGCCCGCTACATCCACCTGCCCGCGGAA 31295
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 400 aacttcctcaactgctgctgcagagcaataagtcggcgagagcaactgcagagacac 459
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31296 CAGATGTTCCCGCTCGCGCCACGACCTCGCGCTGCTGCCGACCTGCCCGCGG 31355
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 460 ctgattggtaccagcgctggtgagccttctgagggcgagccagcatctccg 518
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31356 CTGTACTACGGATCGACCGCGCGCGCGGCTGGAGACCCCTGCCCGGTCAGCG 31414
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
E32986
LOCUS E32986 10732 bp DNA PAT 07-FEB-2001
DEFINITION Gene encoding cellulose synthesizer.
ACCESSION E32986
VERSION E32986.1 GI:13022340
KEYWORDS JP 2000060568-A/1.
SOURCE Vigna angularis.
ORGANISM Vigna angularis.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Vigna.
REFERENCE 1 (bases 1 to 10732)
AUTHORS Koichi, M., T. K. K. and Sato, D. S.
TITLE Gene encoding cellulose synthesizer
JOURNAL Patent: JP 2000060568-A 1 29-FEB-2000;
KOICHI MIZUNO, MITSUJI GIYOUSAI SHOKUBUTSU BIO KENKYUSHO
COMMENT OS Vigna angularis
PN JP 2000060568-A/1
PD 29-FEB-2000
PF 26-AUG-1998 JP 1998239998
PR KOICHI MIZUNO, TOMOHIKO KATO, SHIGERU SATO, DAISUKE SHIBATA PC

```


numbered using the following system eg SC7B7.01c.SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid E29 lies between and overlaps with cosmids E63 and E39 on the AseI-E genomic restriction fragment.

```

FEATURES             source
    1. .26477
    /organism="Streptomyces coelicolor A3(2)"
    /strain="A3(2)"
    /db_xref="taxon:100226"
    /clone="cosmid E29"
    1. .6356
    /gene="SCE63.01"
    /note="complement(1..101)
    <1..6356
    /gene="SCE63.01"
    /note="SCE63.01, partial CDS, cdaPS3, CDA peptide
    synthetase III, len: >332aa; Constitutes the the majority
    of cdaPS3, CDA peptide synthetase III, part of the
    calcium-dependent antibiotic (CDA) biosynthetic cluster
    from Streptomyces coelicolor. CDA is a peptide antibiotic
    which is synthesised non-ribosomally by a putative
    multifunctional peptide synthetase enzyme. This partial
    CDS encodes a subunit of this enzyme suspected to be
    responsible for the addition of 2 amino acids to the
    peptide antibiotic. This ORF overlaps the upstream
    (cdaPSII) by one base indicating possible translational
    coupling. Contains two Pfam matches to entry PF00501
    AMP-binding, AMP-binding enzyme, a Pfam match to entry
    PF00668 DUF4, Domain of unknown function, a Pfam match to
    entry PF00975 Thioesterase. Thioesterase domain and two
    Pfam matches to entry PF00550 pp-binding,
    Phosphopantetheine attachment site. Each PF00501 contains
    a Prosite match PS00455 Putative AMP-binding domain
    signature. Each PF00550 contains a Prosite match PS00012
    Phosphopantetheine attachment site."
    /codon_start=3
    /transl_table=11
    /label=cdaPS3
    /product="CDA peptide synthetase III"
    /protein_id="CAB38876.1"
    /db_xref="GI:4490979"
    /translations="RLRPAETLGLFLRQGEARLIDHQWGLRIORWAGSGELFD
    TAMVFENPLNSRRPGGAAPADLPTLVGRSKDQMHPGLLALPRETLRFSLGI
    LPQVDFPARVEAIAARRALRVLDPADPTRGAVALLDPEVGTLEKWSGDDVRP
    AERTDLFEQVARTFKTALAPDGLTYAELDAANRLARLVELGVGPERHVA
    VGRTELVLAVLAGGAYVVPDPPDRIRHMQADPALVLTSDVDRIEGEE
    CCGPLTFVMDNPTGSLGRHSGTALTADRAAPLLPGHPAVIYTSYTGPRKGVV
    ERALSATFVHRCSQAQDLSGLSVMQASFDQSGSLHAPLISGCVRLTDLRALA
    ETAGSEPGFHRATFMKGTPTSHALLATMPPEVAPSTLTGGEELRGEILAPWREAAG

```

```

DVTVVNYGTEATGCHLEHWIADPTVTEGPPVPICTPHGEGVVRVYVLDLSALRPVAPGL
DCEVYLAGVQLARGYLGRLGGLTAERETDFGAGPSRMYRTGDVAHNEAGELVFAGR
ADQVKLGRYIELGEIEAAVAGGPGVROAAVVRDREDPGDQLVAVYVDPDGHWDEA
AARLALSLPDMFSAFVALDPLSPNGKLDRAALPARTYTGTRAGRAPRTPAE
ILCDLAEVLSLPGVTVDGDFDLGGHSLATLVSRTTTLGAEISIRQFEAPTPA
ALAVLAGAGRAAALTARPRERPLSTAQRLLFWLHLLEGSPYINPTVLRISGP
LRPDALALLDVVGRHESLRTTTEDEGAROVVHPADGVRFVTEADESTEDYSEAD
LARAHAHFDLGAETPVARLLSERHVLILLVHHIADMSRGLAQDUTAAAYTA
RCAGDAPAWQPLFQVADYALMQOETLDDTDPDTLAGLKYWKQQLAGLPLRLDLP
TGDRPATADHTGDRVEFALPADLHRTLTELARATDTTLYMVLQAAALLTLRHGAGE
DIPGTPVAGRTDDATHLVGFVNTLVLTDFSGNPTREDLLTRVGRDLDLTYTHQD
LPERLVEALNPTRSLTHPLFOVLSLSTAPRADGCGAPALPGDRLRVSGTGGAHA
TRAKVDLGSVTERRAADTPDGVAGVLDFTDLDFDRGTAQGLDRLVRVLAADAHP
DRPLSRIDVIGPRERHVEEMNATAGLAPATLPELFEHVRERECAGAYLVADGTSLS
SVAELNARANRLARLLVARGAGPERLVALALPRLALDEEPAAGGEDADLTADRLP
PAIRIAGTDDAAVALLTAAVAAGLPDTPDPRLLDDEEPAAGGEDADLTADRLP
APLLGHPAYVIVTSGTGRKGVTYHSLGALLDIFTSLQDVPGSRVLHLSPAF
DGGFWELMGLLNGAALVVEPGTVPGPALAAALAVRHRVTHAITHAVLQILPEGALP
AGTTLVVAETCPPELVANWSAGRLMRNSYPTTTCATMSAPLAGAAVPIIGRPIA
DGTAYVLDLALQVPVPGVGLYVGGPLARGYLGRPSLTAGRFVACPGAGVVMYR
TGDVLRHADGDLLEYLTDQVKLRGMRVEPAEIEAVTAGLPGVAQAAVLVREDTPG
DRLVGVVVDGASVDPGALROALRGSLPEYMWPAALVVDLALPTTNGKLDHRLP
APEYRTVEGRSPPTPREEALCLFAEVLGELVLDGDFDLGGHSLALRLVERVRA
EELGELGVRDLFAAPTVDLAVRLAARGREPWERLPLRAAGTARPAVPCVHPGSGMS
WCYSGLVRLPLPGIPVYGLQAGLDGDPPLPATLQEMAAEYADLVROTQPEGPYRLIG
WSLGGHVFAMARELARGOEVELLAFDAYPRRAGGAPPAEPAEYFAHNRDAGFDV
AEELTGGRRFPTARYRAFLNAAGDPMGRDLDEAELAAVLEVFMMNAALMRGHTPCTYDG
DLVLAARADGDKLARGAESWRPHVGRGRIERVGDADHGLVQSDAALAVIGRALA
GRDLPATGHUASAAVPETEGTAMNPSPEPAPSPESLSDTEVA"
567. .1794
/gene="SCE63.01"
/note="Pfam match to entry PF00501 AMP-binding,
AMP-binding enzyme, score 385.20, E-value 6.6e-112"
957. .992
/gene="SCE63.01"
/note="PS00455 Putative AMP-binding domain signature."
2046. .2338
/gene="SCE63.01"
/note="Pfam match to entry PF00550 pp-binding,
Phosphopantetheine attachment site, score 78.40, E-value
4.6e-20"
2115. .2162
/gene="SCE63.01"
/note="PS00012 Phosphopantetheine attachment site."
2283. .3480
/gene="SCE63.01"
/note="Pfam match to entry PF00668 DUF4, Domain of unknown
function, score 335.30, E-value 6.9e-97"
3789. .4980
/gene="SCE63.01"
/note="Pfam match to entry PF00501 AMP-binding,
AMP-binding enzyme, score 448.70, E-value 5.2e-131"
4167. .4202
/gene="SCE63.01"
/note="PS00455 Putative AMP-binding domain signature."
5232. .5424
/gene="SCE63.01"
/note="Pfam match to entry PF00550 pp-binding,
Phosphopantetheine attachment site, score 76.90, E-value
1.2e-19"
5301. .5348
/gene="SCE63.01"
/note="PS00012 Phosphopantetheine attachment site."
5490. .6219
/gene="SCE63.01"
/note="Pfam match to entry PF00975 Thioesterase,
Thioesterase domain, score 139.00, E-value 8.5e-38"
6360. .7178
/gene="SCE29.0"
/gene="SCE29.0"
6360. .7178
/gene="SCE29.0"
/note="SCE29.02, probable hydrolase, len: 272aa; similar
to many eg. TR:052809 (EMBL:AJ223998) from the vancomycin

```

```

biosynthesis cluster of Amycolatopsis orientalis (276 aa)
fasta scores: opt: 892, z-score: 1028.9, E(): 0, (50.9%
identity in 271 aa overlap) and TR:O67982 (EMBL:AF003947)
pcal: single polypeptide combining 3'-oxoadipate
enol-lactone hydrolyzing and 4-carboxymuconolactone
decarboxylating activity from Rhodococcus opacus (400 aa)
fasta scores: opt: 288, z-score: 335.4, E(): 2.3e-11,
(32.0% identity in 256 aa overlap). Contains Pfam match to
entry PF00561 abhydrolase, alpha/beta hydrolase fold,
score 77.80, E-value 2.2e-19,"
/codon_start=1
/transl_table=11
/label_SCE29_02
/product="putative hydrolase"
/protein_id="CAB38877.1"
/db_xref="GI:4490980"
/translation="MFVLTNGIRINYDDAPPAGQAQNAFVLVSGSGSGRAWHLH

```

Query Match	1.43;	Score 52.4;	DB 3;	Length 26477;
Best Local Similarity	46.0%;	Pred. No. 0.13;		
Hatches 216;	Conservative 0;	Mismatches 251;	Indels 3;	Gaps 1;
Qy	326	cggggaagtggccagccctggccagcgcgacaccctctgtgacaggcgctggaacac	385	
Db	7425	CGGTGCCCGCCGGTGCCGGTGCCGGCGGCTTCGGCTCCGACCGGAGGGGCGCTGC	7484	
Qy	386	tgacacagccaccacacttctcaactgtatgtctgagagcaataaagtctgcgggagcaga	445	
Db	7485	TGGTCATCGAGCGCTCACGGGAGAGGTGTGGCGCGCGACCGGTTCCCGGTCACTCCGG	7544	
Qy	446	acttgaggagcagctggattggtaaccagcgctgggtgtgagacctcttgaaggcgagc	505	
Db	7545	TCCTCCGGGAGGACTGGCCGGGTGCTGGAGGCGGTGGAGCGCTCGGCACACTGGCGGC	7604	
Qy	506	ccagcatctccggcgccaccataccttcagacccgattcgtgttcgcaccggcccccac	565	
Db	7605	CCGCGCGGCGCGCGCTGGCGGTGTGACTACCGGGGGATGCTGGAGGGGTGCACGCC	7664	
Qy	566	aggtctctctcagggccacgcgcggaggagagcgcatctctgtctccaaagacctgtctct	625	
Db	7665	AGGGCGTCTTCGACGACGGGCACCTGGCGGCACTGTCTCCGGCTCTCCGAACATGTCGGGTG	7724	
Qy	626	ccgcacccccacctggccaacgcacctctggagaccgagtg---gttccagcgctcccgcc	682	
Db	7725	CGCCCCCGAGTTCGGCCACGGGCACCTGGTGTGTGGGGAACGTGTTCCGACGCGGGGCC	7784	
Qy	683	gcaagtggaggcccaacttacaccgcgcggccccaatcaggggcccggggacctggggcc	742	
Db	7785	GGCAGGTCTTGATCGACTGGCGGAGACGCCCTGTGTACTGTCGGGACTGACCTGGGCC	7844	
Qy	743	acagctggcgccgaaggacggtctggcggggacaaagaccacttaag	792	
Db	7845	AGCTCTGGATGCTCTTCGGGAGAGTCCCGGGGCGGGGCGGGCCGCATCAG	7894	

RESULT	15
AK022304	
LOCUS	
DEFINITION	Homo sapiens cDNA FLJ12242 fis, clone MAMMA1001292.
ACCESSION	AK022304
VERSION	AK022304.1 GI:10433671
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens Mammary gland cDNA to mRNA, clone_lib:MAMMA1 clone:MAMMA1001292.
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (sites)	
REFERENCE	
AUTHORS	Isoqai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,N., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,

Search completed: November 1, 2001, 08:59:22
Job time: 8105 sec

TITLE	Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.
REFERENCE	NEDO human cDNA sequencing project
AUTHORS	2 (bases 1 to 1669)
TITLE	Isogai, T. and Otsuki, T.
JOURNAL	Direct Submission
COMMENT	Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812 Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952) NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

```

FEATURES             Location/Qualifiers
     i. 1669
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="MAMMA1001292"
        /clone_lib="MAMMA1"
        /tissue_type="Mammary gland"
        /note="Cloning vector: pME18SFL3"
    2. .874
        /note="unnamed protein product"
        /codon_start=1
        /protein_id="BAB14007.1"
        /db_xref="GI:10433672"
        /translation="MRMEAGEAPGAGGAGGKVVRLNVGGTVLTTRQTLCLR
EQKPSLRQCQGEISDSDRTGAYLIDRPTYFGFLLPLRHGKLVLNDKNAAEEGLV
EEAFYNGILPIIKIADRMEEKDTVTVPHPHVQLQQCEEELTQMVTSTMSDGRRF
EQLVNTGSYSNDSOAEFLFCVWSKELSHTPNGLSSESRKTSTKEGLELEQQQEE
EVVEVEVQVQVEADEAQEKSGRPHLRPEAELAVRASPRPLARPOSCHPCCYKPEAPG
CEAPDHQLGLGVPI"
    309 a          499 c      566 g      295 t
                                     BASE COUNT
                                     ORIGIN

```

Query Match	1.4%	Score 51.8;	DB 89;	Length 1669;
Best Local Similarity	47.6%;	Pred. No. 0.22;		
Matches 152;	Conservative 0;	Mismatches 167;	Indels 0;	Gaps 0;

QY	165	ggaactctcggtcccttggagcgcgtccaccgatggcaccatcttggcgagaaactcgc	224
Db	373	GGACTACAGGTCACCCAGGTGCCACCAACGATGTGTACCGCGTGTCTGCAGTGCCAGGA	432
QY	225	cgaggaggtggccatggacgtggcctcttaccctacacggggactccaccacagctgaa	284
Db	433	GGAGGAGCTCAGGCAATGTCTCCACCATGCTGTGATGGTGGCGCTTCGAGCAGGTGT	492
QY	285	gcgagccaactgtctccggccgcgtacagattggcgggcctgccgggaagtggccagccct	344
Db	493	GAACATCGGTCCTCTTACAACATAGCGCAGCAGGACGAGGAGTCTCTGTGTGTGGT	552
QY	345	ggccagcgcgacaccttccttgacccggcggtggacacactgaacacgcgcccaactt	404
Db	553	GTCCAGGAGGCTCCACAGCACCCCAACGGGCTTGAGTCTCAGAGTCGACGGCGCAAAACCA	612
QY	405	cctcaacgtgatgtcgagagcaataagtgcgcsggagcagaacttcgagagcacctgga	464
Db	513	GAGCACGGAGGACGAGCTGGAGGAGCAGCAGCAGGAGGAGGAGGTGGAGGAGGTGGA	672
QY	165	ttggtaccaggcgctgggt	483
Db	573	GGTGAACAGGTGCAGGTG	691

Search completed: November 1, 2001, 08:59:22
Job time: 8105 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2001, 06:42:12 ; Search time 159.69 Seconds
(without alignments)
14343.948 Million cell updates/sec

Title: US-09-775-181-1

Perfect score: 3648

Sequence: 1 atggagcattgcttaacc.....gggtagttttaaaagttag 3648

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601:*

- 1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	342.2	9.4	368	21	AAA42817 Human secreted exp
2	156.8	4.3	182	21	AAA41583 Human secreted exp
3	65.4	1.8	936	22	AAF58252 Oligonucleotide D1
4	65.4	1.8	936	22	AAF58254 Oligonucleotide D1
5	65.4	1.8	936	22	AAF58257 Oligonucleotide D1
6	65.4	1.8	936	22	AAF58259 Oligonucleotide D2
7	65.4	1.8	936	22	AAF58262 Oligonucleotide D2
8	65.4	1.8	936	22	AAF58255 Oligonucleotide D1
9	64	1.8	83	21	AAC08885 Human secreted pro
10	63.2	1.7	936	22	AAF58252 Oligonucleotide D1
11	63.2	1.7	936	22	AAF58254 Oligonucleotide D1

c 12	63.2	1.7	936	22	AAF58257	Oligonucleotide D1
c 13	63.2	1.7	936	22	AAF58259	Oligonucleotide D2
c 14	63.2	1.7	936	22	AAF58262	Oligonucleotide D1
c 15	63.2	1.7	938	22	AAF58255	Oligonucleotide D1
c 16	56.4	1.5	10732	21	AAA10594	Gene encoding a su
c 17	52	1.4	1035	21	AAA51616	HIV gp41 coding re
c 18	52	1.4	2466	21	AAA51614	HIV gp160 coding r
c 19	52	1.4	2547	21	AAA51615	HIV gp160 with sig
c 20	51.2	1.4	1459	21	AAC69799	Human breast tumou
c 21	51.2	1.4	3867	22	AAF28253	Human TANGO 275 op
c 22	51.2	1.4	4225	22	AAF28252	Human TANGO 275 DN
c 23	48	1.3	2571	20	AAZ08742	HIV SYNgp-160mm co
c 24	48	1.3	2571	21	AAZ93974	Human immunodefici
c 25	48	1.3	2571	22	AAC86878	Nucleotide sequenc
c 26	46.4	1.3	2481	17	AAZ13557	Syngp160mm. Synth
c 27	46.4	1.3	2481	18	AAZ73952	HIV-1 gp160 synthe
c 28	46.4	1.3	2481	19	AAZ23291	Synthetic HIV-1 gp
c 29	45.6	1.2	2298	21	AAA51963	Modified HIV-1 Env
c 30	45.6	1.2	2298	21	AAA51964	Modified HIV-1 Env
c 31	45.6	1.2	2298	21	AAA51965	Modified HIV-1 Env
c 32	45.6	1.2	2310	21	AAA51944	Modified HIV-1 Env
c 33	45.6	1.2	2310	21	AAA51962	Modified HIV-1 Env
c 34	45.6	1.2	2316	21	AAA51945	Modified HIV-1 Env
c 35	45.6	1.2	2316	21	AAA51949	Modified HIV-1 Env
c 36	45.6	1.2	2322	21	AAA51946	Modified HIV-1 Env
c 37	45.6	1.2	2322	21	AAA51959	Modified HIV-1 Env
c 38	45.6	1.2	2322	21	AAA51959	Modified HIV-1 Env
c 39	45.6	1.2	2322	21	AAA51961	Modified HIV-1 Env
c 40	45.6	1.2	2328	21	AAA51947	Modified HIV-1 Env
c 41	45.6	1.2	2328	21	AAA51948	Modified HIV-1 Env
c 42	45.6	1.2	2352	21	AAA51967	Modified HIV-1 Env
c 43	45.6	1.2	2358	21	AAZ70447	HIV gp160.modSf162
c 44	45.6	1.2	2358	21	AAA51966	Modified HIV-1 Env
c 45	45.6	1.2	2466	21	AAA70446	HIV gp160.modSf162

ALIGNMENTS

RESULT 1
AAA42817
ID AAA42817 standard; cDNA; 368 BP.
XX
AC AAA42817;
XX
XX 21-AUG-2000 (first entry)
XX
DE Human secreted expressed sequence tag SEQ ID NO:1557.
XX

Human; mouse; xenopus; rat; secreted expressed sequence tag; sEST;
expressed sequence tag; EST; probe; chemotactic; proliferative;
immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
chemilucifer; osteopathic; neuroprotective; nootropic; antipsoriatic;
vaccine; autoimmune disorder; anticonvulsant; antidepressant; gene therapy;
insulin dependent diabetes; multiple sclerosis; allergic condition;
lymphoid cell deficiency; asthma; myeloid cell deficiency; ulcer;
central nervous system disorder; osteoarthritis;
Parkinson's disease; Huntington's disease; coagulation disorder;
thrombophilia; thrombosis; inflammatory disorder; Crohn's disease;
cancer; infection; depression; psoriasis; ss.

Human sapiens.

W2000021990-A1.

20-APR-2000.

15-OCT-1999; 99WO-US24205.

15-OCT-1998; 98US-0104435.

```
XX (GEMY ) GENETICS INST INC.
XX
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M;
XX WPI; 2000-317937/27.
XX
XX Isolated polynucleotides, and encoded proteins, comprising secreted
XX expressed sequence tags (sESTs), useful for treating various disorders
XX such as autoimmune, infectious, and central nervous system disorders -
XX
XX Claim 1; Page 487; 618pp; English.
XX
XX AAA41261 to AAA43419 represent specifically claimed secreted expressed
XX sequence tags (sESTs), isolated from human, mouse, xenopus and rat
XX tissue sources. The sESTs can have a range of activities depending on
XX the tissues they were isolated from. The activities include:
XX chemotactic; proliferative; immunomodulatory; haematopoietic;
XX chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
XX cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
XX antiasthmatic; vulnery; antitumor; osteopathic; neuroprotective;
XX neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective;
XX anticonvulsant; and antidepressant. The sESTs can be used for gene
XX therapy and in vaccines. The sESTs are useful as probes for the
XX identification and isolation of full-length cDNAs and genomic DNA
XX molecules which correspond to the sESTs. Proteins encoded by the sESTs
XX are useful in assays for determining biological activity and raising
XX antibodies. They may be useful for treatment of autoimmune disorders
XX (multiple sclerosis, insulin dependent diabetes), allergic conditions
XX (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
XX osteoporosis, osteoarthritis, central nervous system disorders
XX (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
XX disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
XX disease), tumours, bacterial, fungal or viral infections, depression and
XX psoriasis. AAA43420 to AAA43425 represent linker variants which are given
XX in the exemplification of the present invention.
XX
XX Sequence 368 BP; 115 A; 78 C; 79 G; 96 T; 0 other;
XX
XX
XX Query Match 9.4%; Score 342.2; DB 21; Length 368;
XX Best Local Similarity 99.1%; Pred. No. 1e-83;
XX Matches 344; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 1871 ctgtatattccatacaattagattgttcttgcctcaagacttcagtcgtattggtatg 1930
XX || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 19 ctactatattccatacaattagattgttcttgcctcaagacttcagtcgtattggtatg 78
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 1931 tgaatgctgtatttgcacatactcattgactgtgacagtcaccattgggtgcttttga 1990
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 79 tgaatgctgtatttgcacatactcattgactgtgacagtcaccattgggtgcttttga 138
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 1991 ttccaaagtattccatacaattcaagcaataccacagagatgattgtcacaagatagtg 2050
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 139 ttccaaagtattccatacaattcaagcaataccacagagatgattgtcacaagatagtg 198
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 2051 aggatgagtagacatgggcccagctggtatcctcactgaacagcagatcaattcagcct 2110
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 199 aggatgagtagacatgggcccagctggtatcctcactgaacagcagatcaattcagcct 258
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 2111 ggagttagcacagcttggtaccagagacattcgggacagagctgaaataactctatgcc 2170
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 259 ggagttagcacagcttggtaccagagacattcgggacagagctgaaataactctatgcc 318
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 2171 aactggaaatataaagaagaagatgattacaaaacacccccac 2217
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 319 aactggaaatataaagaagaagatgattacaaaacacccccac 365
```

RESULT 2

AAA41583

ID AAA41583 standard; cDNA; 182 BP.

```
XX AAA41583;
XX
XX 21-AUG-2000 (first entry)
XX
XX Human secreted expressed sequence tag SEQ ID NO:323.
XX
XX Human; mouse; xenopus; rat; secreted expressed sequence tag; sEST;
XX expressed sequence tag; EST; probe; chemotactic; proliferative;
XX immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
XX thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
XX antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;
XX cerebroprotective; anticonvulsant; antidepressant; gene therapy;
XX vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
XX central nervous system disorder; Alzheimer's disease; stroke;
XX Parkinson's disease; Huntington's disease; coagulation disorder;
XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
XX tumour; infection; depression; psoriasis; ss.
XX
XX Homo sapiens.
XX
XX WO2000021990-A1.
XX
XX 20-APR-2000.
XX
XX 15-OCT-1999; 99WO-US24205.
XX
XX 15-OCT-1998; 98US-0104435.
XX
XX (GEMY ) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M;
XX WPI; 2000-317937/27.
XX
XX Isolated polynucleotides, and encoded proteins, comprising secreted
XX expressed sequence tags (sESTs), useful for treating various disorders
XX such as autoimmune, infectious, and central nervous system disorders -
XX
XX Claim 1; Page 243; 618pp; English.
XX
XX AAA41261 to AAA43419 represent specifically claimed secreted expressed
XX sequence tags (sESTs), isolated from human, mouse, xenopus and rat
XX tissue sources. The sESTs can have a range of activities depending on
XX the tissues they were isolated from. The activities include:
XX chemotactic; proliferative; immunomodulatory; haematopoietic;
XX chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
XX cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
XX antiasthmatic; vulnery; antitumor; osteopathic; neuroprotective;
XX neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective;
XX anticonvulsant; and antidepressant. The sESTs can be used for gene
XX therapy and in vaccines. The sESTs are useful as probes for the
XX identification and isolation of full-length cDNAs and genomic DNA
XX molecules which correspond to the sESTs. Proteins encoded by the sESTs
XX are useful in assays for determining biological activity and raising
XX antibodies. They may be useful for treatment of autoimmune disorders
XX (multiple sclerosis, insulin dependent diabetes), allergic conditions
XX (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
XX osteoporosis, osteoarthritis, central nervous system disorders
XX (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
XX disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
XX disease), tumours, bacterial, fungal or viral infections, depression and
XX psoriasis. AAA43420 to AAA43425 represent linker variants which are given
XX in the exemplification of the present invention.
XX
XX Sequence 182 BP; 46 A; 36 C; 49 G; 51 T; 0 other;
XX
XX
XX Query Match 4.3%; Score 156.8; DB 21; Length 182;
```

Best Local Similarity 93.2%; Pred. No. 3.9e-33;

Matches 164; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1529 ttcttccacgaacggtcgaacgaattccatatatgactggtgacggtgtcatgagatgc 1588

Db 4 ttcggccaaagagcctaattgaattccatatatgactggtgacggtgtcatgagatgc 63

QY 1589 tgcagtaataactcttggttagtgtttgtttctcatctggtggaactctctgtgtgcc 1648

Db 64 tgcagtaataactcttggttagtgtttgtttctcatctggtggaactctctgtgtgcc 123

QY 1649 agaattggagaacagattctacttatgtccaggaggaacacatccgataccctc 1704

Db 124 agaattggagaacagattctacttatgtccaggaggaacacacccgataccctc 179

RESULT 3

AAF58252
ID AAF58252 standard; DNA; 936 BP.

AAF58252;

24-APR-2001 (first entry)

Oligonucleotide D1835.

Electron-transfer group; ETM; mismatch; genotyping;
gene expression; ss.

Synthetic.

WO200107665-A2.

01-FEB-2001.

26-JUL-2000; 2000WO-US20476.

26-JUL-1999; 99US-0145695.

17-MAR-2000; 2000US-0190259.

(CLIN-) CLINICAL MICRO SENSORS INC.

Umek RM;

WPI; 2001-159728/16.

Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface.

Example 6; Page 127; 159pp; English.

The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, and monitoring gene expression.

Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 1.8%; Score 65.4; DB 22; Length 936;

Best Local Similarity 2.2%; Pred. No. 1.1e-07;

Matches 18; Conservative 436; Mismatches 357; Indels 0; Gaps 0;

QY 1411 attttgactttgagccaaacacattctgctgtattctcctaagatggctgtctctc 1470

Db 1 www. 60

QY 1471 ggtttgtactgttttacggaactgtcactctcaaaacttcacagggttttgaagggtttt 1530

Db 61 www. 120

QY 1531 ctttcaacgaacggtcgaacgaattccatatatgactggtgacggtgtcatgagatgctg 1590

Db 121 www. 180

QY 1591 gcagtaataactcttggttagtgtttgtttctcatctggtggaacttcattctgtgtgccag 1650

Db 181 www. 240

QY 1651 aatttgagaacagattctacttatgtccagggtggaacacatccgataccctcattc 1710

Db 241 www. 300

QY 1711 aatatgtccattgacccgtggtgactacatgacagcagttgtgtaattttattctc 1770

Db 301 www. 360

QY 1771 ttgtgggtgttttctctctctatgcagtggtgacagtcctccatcggtcattccatgagccc 1830

Db 361 www. 420

QY 1831 cgctatatggtgtgtgagttcacaatgagtcattctctctgtattccatacaatt 1890

Db 421 www. 480

QY 1891 agatttggcttgcctcaagacttcagtcgtgattggtgattgtgtgtattttgcacat 1950

Db 481 www. 540

QY 1951 acatcatttgactgtgacagtcacattgggttgcctttgattccaaagttttcaatca 2010

Db 541 www. 600

QY 2011 agcaataaccacagagatgatattgtcacagaagcatatgagatgagctagacatgggc 2070

Db 601 www. 660

QY 2071 cgatctggatcctacctgaacagcagtcataatcagcctggagtgagcacagcttggat 2130

Db 661 www. 720

QY 2131 ccagaggacattcgggacgagctgaaaaaactctctatcccaactggaataataaaaga 2190

Db 721 www. 780

QY 2191 aagaagatgatcacaaacaccccccacctcc 2221

Db 781 www. 811

RESULT 4

AAF58254

ID AAF58254 standard; DNA; 936 BP.

XX AAF58254;

AC AAF58254;

DT 24-APR-2001 (first entry)

XX Oligonucleotide D1875.

XX Electron-transfer group; ETM; mismatch; genotyping;

XX gene expression; ss.

XX Synthetic.

XX WO200107665-A2.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-US20476.

XX 26-JUL-1999; 99US-0145695.

XX 17-MAR-2000; 2000US-0190259.


```
DT 24-APR-2001 (first entry)
XX Oligonucleotide D1835.
XX Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX Synthetic.
OS WO200107665-A2.
PN 01-FEB-2001.
PD 26-JUL-2000; 2000WO-US20476.
XX 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX (CLIN-) CLINICAL MICRO SENSORS INC.
PA Umek RM;
XX WPI; 2001-159728/16.
DR Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface.
PS Example 6; Page 127; 159pp; English.
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;
S0

Query Match 1.7%; Score 63.2; DB 22; Length 936;
Best Local Similarity 1.1%; Pred. No. 4.4e-07;
Matches 8; Conservative 411; Mismatches 319; Indels 0; Gaps 0;

Qy 1473 ttctgactgtttacggaactgtcactctcaactcacagggttttgaagggtttctt 1532
Db 759 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 700

1533 ttcacgaacggtcacaagaattccatatgactgacgagcgggtcatgagatgctgc 1592
699 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 640

Qy 1593 agtaatactctgttagtgtttgtttctctattggtgacttcattctgtgccaaga 1652
Db 639 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 580

Qy 1653 ttgtgagaaacagatttccatttattggccagggaacacccgatccctcattcaa 1712
Db 579 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 520

Qy 1713 tatgtcctcattgacgcctggactacatgacagcagtgctgacttattctcttt 1772
Db 519 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 460

Qy 1773 gtgggggttttactctgctatgcagtcgacagtcgccatccatccatgagcccg 1832
Db 459 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 400

Qy 1833 ctatatggtgttgctgagttcacatgagctcatctctctctatattccatacaattag 1892
Db 399 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 340

Qy 1893 attgttcttctcctcaagacttcagtcgtgattgagtgctgtgtgttttgcacatac 1952
```

```
Db 339 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 280
Qy 1953 tcatttgactgtgacagtcaccattgggttgctttgtattccaaagttttcacattcaag 2012
Db 279 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 220
Qy 2013 caataaccacagatgatattgtctacagaagcatatgagtgatgagctagacatggccg 2072
Db 119 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 160
Qy 2073 atctggtactctacctaagcagcagtcattcaattcagcctgagtgacacagcttgatcc 2132
Db 159 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 100
Qy 2133 agagacattcgggacgagctgaaacaaactctatcccaactggaataataaaagaaa 2192
Db 99 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 40
Qy 2193 gaagatgatcacacaaa 2210
Db 39 wwwwwwwwwwwwwwwwwwwww 22

RESULT 11
AAF58254/c
ID AAF58254 standard; DNA; 936 BP.
XX AC AAF58254;
XX DT 24-APR-2001 (first entry)
XX DE Oligonucleotide D1875.
XX KW Electron-transfer group; ETM; mismatch; genotyping;
XX KW gene expression; ss.
XX OS Synthetic.
XX PN WO200107665-A2.
XX PD 01-FEB-2001.
XX PF 26-JUL-2000; 2000WO-US20476.
XX PR 26-JUL-1999; 99US-0145695.
XX PR 17-MAR-2000; 2000US-0190259.
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI Umek RM;
XX WPI; 2001-159728/16.
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface.
XX Example 6; Page 127; 159pp; English.
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;
S0

Query Match 1.7%; Score 63.2; DB 22; Length 936;
Best Local Similarity 1.1%; Pred. No. 4.4e-07;
Matches 8; Conservative 411; Mismatches 319; Indels 0; Gaps 0;
```


Search completed: November 1, 2001, 07:47:23
Job time: 3911 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2001, 07:46:12 ; Search time 66.1 Seconds
(without alignments)
10447.912 Million cell updates/sec

Title: US-09-775-181-1
Perfect score: 3648
Sequence: 1 atggagcagtgcttacc.....gggtagtatttaagtagttag 3648
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A-COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B-COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A-COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B-COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS-COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	62	1.7	7218	1	US-08-232-463-14
2	48.4	1.3	7218	1	US-08-232-463-14
3	46.4	1.3	2481	1	US-08-324-243-35
4	46.4	1.3	2481	1	US-08-532-390-35
5	46.4	1.3	2481	1	US-08-717-294-35
6	46.4	1.3	2481	5	PCT-US95-11511-35
7	44.8	1.2	2502	1	US-08-073-384C-7
8	44.8	1.2	2502	1	US-08-254-359A-7
9	44.8	1.2	2502	1	US-08-483-043-7
10	44.8	1.2	2502	1	US-08-481-238-7
11	44.8	1.2	2502	2	US-08-471-066B-7
12	44.8	1.2	2502	2	US-08-484-956-7
13	44.8	1.2	2502	2	US-08-757-653-7
14	44.8	1.2	2502	2	US-08-599-491-7
15	44.8	1.2	2502	2	US-08-756-386-7
16	44.8	1.2	2502	2	US-08-823-516-7
17	44.8	1.2	2502	3	US-08-682-853A-7
18	44.8	1.2	2502	3	US-08-759-038-7
19	44.8	1.2	2502	3	US-08-758-314-7
20	44.8	1.2	2634	3	US-08-911-853-30
21	44.8	1.2	2634	3	US-08-479-409-30
22	44.8	1.2	17612	3	US-08-911-853-29
23	44.8	1.2	17612	4	US-09-479-409-29
24	42.8	1.2	771	2	US-08-972-008-3
c 25	42.8	1.2	2525	2	US-08-972-008-1
26	42	1.2	44377	2	US-08-804-227C-7
27	42	1.2	44377	2	US-08-804-198-1

29	41.8	1.1	2496	1	US-08-073-384C-2	Sequence 2, Appli
29	41.8	1.1	2496	1	US-08-254-359A-2	Sequence 2, Appli
30	41.8	1.1	2496	1	US-08-483-043-2	Sequence 2, Appli
31	41.8	1.1	2496	1	US-08-481-238-2	Sequence 2, Appli
32	41.8	1.1	2496	1	US-08-471-066B-2	Sequence 2, Appli
33	41.8	1.1	2496	2	US-08-484-956-2	Sequence 2, Appli
34	41.8	1.1	2496	2	US-08-757-653-2	Sequence 2, Appli
35	41.8	1.1	2496	2	US-08-599-491-2	Sequence 2, Appli
36	41.8	1.1	2496	2	US-08-756-386-2	Sequence 2, Appli
37	41.8	1.1	2496	2	US-08-823-516-2	Sequence 2, Appli
38	41.8	1.1	2496	3	US-08-682-853A-2	Sequence 2, Appli
39	41.8	1.1	2496	3	US-08-759-038-2	Sequence 2, Appli
40	41.8	1.1	2496	3	US-08-758-314-2	Sequence 2, Appli
41	40.6	1.1	4776	2	US-08-852-401-1	Sequence 1, Appli
c 42	40.4	1.1	12001	1	US-08-458-568A-11	Sequence 11, Appl
43	40.4	1.1	1524	2	US-08-716-942-24	Sequence 24, Appl
c 44	40.4	1.1	1771	2	US-08-533-669A-7	Sequence 7, Appli
c 45	40.4	1.1	1771	2	US-08-511-872-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzgpt-Fls
; US-08-232-463-14

CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 04-JUN-1993
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/986,330
FILING DATE: 07-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-00613
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/705-8410
TELEFAX: 415/397-8338
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2502 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-073-384C-7

Query Match 1.2%; Score 44.8; DB 1; Length 2502;
Best Local Similarity 46.8%; Pred. No. 0.038;
Matches 176; Conservative 0; Mismatches 197; Indels 3; Gaps 1;
Qy 413 tgatgctgcagagcaataagtcgaggagcagaacttcgaggagcactgattggtacc 472
Db 1820 TGGTGGCCCTGGACTATAGCCAGATAGAGCTCCGGGTCTTGGCCACCTCTCCGGGGACG 1879
Qy 473 aggcgtggtgtgagccttctgagggcgagcccgagcactctccggcgccacacct 532
Db 1880 AGACCTGATCCGGGTCTTCCAGGAGGGGAGGACATCCACACCCAGCCAGCTGGA 1939
Qy 533 tcagcacgattcgtctccgacccgagggcaggggtcttctccagggcagcgag 592
Db 1940 TGTTCCGGCGTCCCGCGAGCCGTGACCCCTGATGCGCGCGCGCCCAAGACCATCA 1999
Qy 593 agagccgacatctgctccagacactgctctctccgacccacacacactggccacg 652
Db 2000 ACTTCGGGGTC---CTCTACGGCATGTCCGCCACCCACCCCTCTCCAGGAGCTTCCCATCC 2056
Qy 653 tggagaccgagtggttccacggcctccgagggcgagggagggagggcacttacacgcgcg 712
Db 2057 CCTACGAGGAGCGGTGGCTTCTTCCAGGCGCTACTTCCAGAGCTTCCCAAGTGGCGG 2116
Qy 713 gccccaatcagggggcccggggcttggccacagctggcgcgcaagaggggctcgcg 772
Db 2117 CCTGATTGAGAACCCCTTGAGGAGGCGAGGGCGGGGTACGTGGAGACCCTCTTCG 2176
Qy 773 gggacaagaccactt 788
Db 2177 GCCCGCGCGCTACGT 2192

RESULT 8
US-08-254-359A-7
Sequence 7, Application US/08254359A
Patent No. 5614402
GENERAL INFORMATION:
APPLICANT: DAHLBERG, JAMES E.
APPLICANT: LYANICHEV, VICTOR I.
APPLICANT: BROW, MARY ANN D.

TITLE OF INVENTION: 5' NUCLEASES DERIVED FROM THERMOSTABLE
TITLE OF INVENTION: DNA POLYMERASE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/254,359A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/073,384
FILING DATE: 06-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/986,330
FILING DATE: 07-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2502 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-254-359A-7

Query Match 1.2%; Score 44.8; DB 1; Length 2502;
Best Local Similarity 46.8%; Pred. No. 0.038;
Matches 176; Conservative 0; Mismatches 197; Indels 3; Gaps 1;
Qy 413 tgatgctgcagagcaataagtcgaggagcagaacttcgaggagcactgattggtacc 472
Db 1820 TGGTGGCCCTGGACTATAGCCAGATAGAGCTCCGGGTCTTGGCCACCTCTCCGGGGACG 1879
Qy 473 aggcgtggtgtgagccttctgagggcgagcccgagcactctccggcgccacacct 532
Db 1880 AGACCTGATCCGGGTCTTCCAGGAGGGGAGGACATCCACACCCAGCCAGCTGGA 1939
Qy 533 tcagcacgattcgtctccgacccgagggcaggggtcttctccagggcagcgag 592
Db 1940 TGTTCCGGCGTCCCGCGAGCCGTGACCCCTGATGCGCGCGCGCCCAAGACCATCA 1999
Qy 593 agagccgacatctgctccagacactgctctctccgacccacacacactggccacg 652
Db 2000 ACTTCGGGGTC---CTCTACGGCATGTCCGCCACCCACCCCTCTCCAGGAGCTTCCCATCC 2056
Qy 653 tggagaccgagtggttccacggcctccgagggcgagggcacttacacgcgcgcg 712
Db 2057 CCTACGAGGAGCGGTGGCTTCTTCCAGGCGCTACTTCCAGAGCTTCCCAAGTGGCGG 2116
Qy 713 gccccaatcagggggcccggggcttggccacagctggcgcgcaagaggggctcgcg 772
Db 2117 CCTGATTGAGAACCCCTTGAGGAGGCGAGGGCGGGGTACGTGGAGACCCTCTTCG 2176
Qy 773 gggacaagaccactt 788
Db 2177 GCCCGCGCGCTACGT 2192

RESULT 9
US-08-483-043-7
; Sequence 7, Application US/08483043
; Patent No. 5691142
; GENERAL INFORMATION:
; APPLICANT: Dahlberg, James E.
; APPLICANT: Lyamichiev, Victor I.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
; TITLE OF INVENTION: POLYMERASE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483.043
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/073,384
; FILING DATE: 04-JUN-1993
; APPLICATION NUMBER: US 07/986,330
; FILING DATE: 07-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Cartoli, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-00613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/705-8410
; TELEFAX: 415/397-8338
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-483-043-7

Query Match 1.2%; Score 44.8; DB 1; Length 2502;
Best Local Similarity 46.8%; Pred. No. 0.038;
Matches 176; Conservative 0; Mismatches 197; Indels 3; Gaps 1;
Qy 413 tgatgtcgagagcaataagtcgaggagagcaacttcgaggagacacctggaattgtacc 472
Db 1820 TGGTGGCCCTGGACTATAGACAGATAGAGTCCGGGTCTCTGGCCACCTCTCTCCGGGACG 1879
Qy 473 aggcgtggttgagccttctgaggcgagccagcatctcccgggcgccatcacct 532
Db 1880 AGAACCCTGATCCGGGTCTTCCAGGAGGGAGGAGCATCCACACAGACCCGACCTCGGA 1939
Qy 533 tcagcaccgattcgctgctccgacccgcccacaggttctctccagccagcgccgag 592
Db 1940 TGTTCGGCGTCCCCCGGAGGCCGTGGACCCCTGATGCGCGGGCGGCGGACCATCA 1999
Qy 593 agagccgcatctgctccaagacctgtctctctccgacccacacacacacacacacac 652
Db 2000 ACTTCGGGGTC---CTCTACGGCATGTCCGCCACCGCTCTCTCCAGGAGCTTGCCATCC 2056
Qy 653 tggagaccgagtggttccacggcctccggcgcaagtggagggcccccacattacacccgc 712
Db 2057 CCTACGAGGAGCGGTGGCCCTTCATTGAGCGCTACTTCCAGAGCTTCCCCAAGGTGCGG 2116

Qy 713 gcccaatcagggcccccgggacctgaggccacagctggcgcgcaaggagcggtcgcg 772
Db 2117 CCTGATTGAGAGACCCCTGGAGAGGGGAGGAGCGGGGGTACGTGGAGACCCCTCTTCG 2176
Qy 773 gggacaagagccactt 788
Db 2177 GCCGCGGGCTACGT 2192

RESULT 10
US-08-481-238-7
; Sequence 7, Application US/08481238
; Patent No. 5795763
; GENERAL INFORMATION:
; APPLICANT: DAHLBERG, JAMES E.
; APPLICANT: LYAMICHEV, VICTOR I.
; APPLICANT: BROW, MARY ANN D.
; TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
; TITLE OF INVENTION: POLYMERASE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,238
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-01798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-481-238-7

Query Match 1.2%; Score 44.8; DB 1; Length 2502;
Best Local Similarity 46.8%; Pred. No. 0.038;
Matches 176; Conservative 0; Mismatches 197; Indels 3; Gaps 1;
Qy 413 tgatgtcgagagcaataagtcgaggagagcaacttcgaggagacacctggaattgtacc 472
Db 1820 TGGTGGCCCTGGACTATAGACAGATAGAGTCCGGGTCTCTGGCCACCTCTCTCCGGGACG 1879
Qy 473 aggcgtggttgagccttctgaggcgagccagcatctcccgggcgccatcacct 532
Db 1880 AGAACCCTGATCCGGGTCTTCCAGGAGGGAGGAGCATCCACACAGACCCGACCTCGGA 1939
Qy 533 tcagcaccgattcgctgctccgacccgcccacaggttctctccagccagcgccgag 592
Db 1940 TGTTCGGCGTCCCCCGGAGGCCGTGGACCCCTGATGCGCGGGCGGCGGACCATCA 1999
Qy 593 agagccgcatctgctccaagacctgtctctctccgacccacacacacacacacacac 652
Db 2000 ACTTCGGGGTC---CTCTACGGCATGTCCGCCACCGCTCTCTCCAGGAGCTTGCCATCC 2056


```
; REFERENCE/DOCKET NUMBER: FORS-01801
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-484-956-7

Query Match          1.2%; Score 44.8; DB 2; Length 2502;
Best Local Similarity 46.8%; Pred. No. 0.038;
Matches 176; Conservative 0; Mismatches 197; Indels 3; Gaps 1;

QY 413 tgatgctgcagagcaataagtcgcgggagcagaacttgcagagcacttgattgtgacc 472
   || || || || || || || || || || || || || || || || || || || || ||
1820 TGGTGGCCCTGGACTATAGCCAGATAGAGCTCCGGGTCTCTGGCCCACTCTCCGGGGCAG 1879

QY 473 aggcgtgtgtgagccttcttgaggcgagcccaagcatctccggggggccatcacct 532
   || || || || || || || || || || || || || || || || || || || || ||
Db 1880 AGAACCTGATCCGGGTCTTCCAGGAGGGAGGACATCCACACCCAGCCGACGCTGGA 1939

QY 533 tcagcacccattcgtctgcgcacgcggcccccacaggtctctccagggccgcgcgag 592
   || || || || || || || || || || || || || || || || || || || || ||
Db 1940 TGTTCCGGCTCCCGGAGCGGTGGACCCCTGATGCGCGGGGGCCAGACCATCA 1999

QY 593 agagcgcacatctgtctccaaagacctgtctctctcgcaccccaacctggccacgcactc 652
   || || || || || || || || || || || || || || || || || || || || ||
Db 2000 ACTTCCGGGTC---CTCTACGGCATCTCCGCCACCGCCCTCTCCAGGAGCTTGCCATCC 2056

QY 653 tggagaccagtggttccacgcctccgcgcgaagtggaggcccaacttaaccgcgcg 712
   || || || || || || || || || || || || || || || || || || || || ||
Db 2057 CCTACGAGAGGGGTGGCGCTTCTATGTAGCGCTACTTCCAGAGCTTCCCAAGGTGCGGG 2116

QY 713 gccccaatcagggggcccccggcctgggcccacagctggcgcgcaaggagcggtctcg 772
   || || || || || || || || || || || || || || || || || || || || ||
Db 2117 CCTGGATTGAGAAGACCCCTGGAGGAGGGGAGGCGGGGGTACGTGGAGACCTCTTCG 2176

QY 773 gggacaagagccactt 788
   || || || || || || || || || || || || || || || || || || || || ||
Db 2177 GCCCGCGCGCTACGT 2192
```

RESULT 13

```
; US-08-757-653-7
; Sequence 7, Application US/08757653
; Patent No. 5843669
; GENERAL INFORMATION:
; APPLICANT: Kaiser, Michael W.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Lyamichev, Natasha
; TITLE OF INVENTION: Cleavage Of Nucleic Acid Using
; TITLE OF INVENTION: Thermostable FEN-1 Endonucleases
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,653
; FILING DATE:
```

```
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-02565
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-757-653-7

Query Match          1.2%; Score 44.8; DB 2; Length 2502;
Best Local Similarity 46.8%; Pred. No. 0.038;
Matches 176; Conservative 0; Mismatches 197; Indels 3; Gaps 1;

QY 413 tgatgctgcagagcaataagtcgcgggagcagaacttgcagagcacttgattgtgacc 472
   || || || || || || || || || || || || || || || || || || || || ||
Db 1820 TGGTGGCCCTGGACTATAGCCAGATAGAGCTCCGGGTCTCTGGCCCACTCTCCGGGGCAG 1879

QY 473 aggcgtgtgtgagccttcttgaggcgagcccaagcatctccggggggccatcacct 532
   || || || || || || || || || || || || || || || || || || || || ||
Db 1880 AGAACCTGATCCGGGTCTTCCAGGAGGGAGGACATCCACACCCAGCCGACGCTGGA 1939

QY 533 tcagcacccattcgtctgcgcacgcggcccccacaggtctctccagggccgcgcgag 592
   || || || || || || || || || || || || || || || || || || || || ||
Db 1940 TGTTCCGGCTCCCGGAGCGGTGGACCCCTGATGCGCGGGGGCCAGACCATCA 1999

QY 593 agagcgcacatctgtctccaaagacctgtctctctcgcaccccaacctggccacgcactc 652
   || || || || || || || || || || || || || || || || || || || || ||
Db 2000 ACTTCCGGGTC---CTCTACGGCATCTCCGCCACCGCCCTCTCCAGGAGCTTGCCATCC 2056

QY 653 tggagaccagtggttccacgcctccgcgcgaagtggaggcccaacttaaccgcgcg 712
   || || || || || || || || || || || || || || || || || || || || ||
Db 2057 CCTACGAGAGGGGTGGCGCTTCTATGTAGCGCTACTTCCAGAGCTTCCCAAGGTGCGGG 2116

QY 713 gccccaatcagggggcccccggcctgggcccacagctggcgcgcaaggagcggtctcg 772
   || || || || || || || || || || || || || || || || || || || || ||
Db 2117 CCTGGATTGAGAAGACCCCTGGAGGAGGGGAGGCGGGGGTACGTGGAGACCTCTTCG 2176

QY 773 gggacaagagccactt 788
   || || || || || || || || || || || || || || || || || || || || ||
Db 2177 GCCCGCGCGCTACGT 2192
```

RESULT 14

```
; US-08-599-491-7
; Sequence 7, Application US/08599491
; Patent No. 5846717
; GENERAL INFORMATION:
; APPLICANT: BROW, MARY ANN D.
; APPLICANT: GROTELUESCHEN HALL, JEFF S.
; APPLICANT: LYAMICHEV, VICTOR
; APPLICANT: OLIVE, DAVID M.
; APPLICANT: PRUDENT, JAMES R.
; TITLE OF INVENTION: DETECTION OF NUCLEIC ACID SEQUENCES BY
; TITLE OF INVENTION: INWARD-DIRECTED CLEAVAGE
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```


COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/599,491
FILING DATE: 23-JAN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: P-40,027
REFERENCE/DOCKET NUMBER: FORS-01802
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2502 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-599-491-7

Query Match 1.2%; Score 44.8; DB 2; Length 2502;
Best Local Similarity 46.8%; Pred. No. 0.038;
Matches 176; Conservative 0; Mismatches 197; Indels 3; Gaps 1;

QY 413 tgatgctgcagagcaataagtcgaggagcagaacttgcaggacgacctggtattgtacc 472
DB 1820 TGGTGGCCCTGGACTATAGCCAGATAGAGTCCGGGTCTTGGCCACCTCTCCGGGAGC 1879
QY 473 aggcgtggtgtgagccttcttgaggcgcagccagcatctccgggcgcctacacct 532
DB 1880 AGAACCTGTATCCGGGTCTTCCAGGAGGGAGGACATCCACACCCAGCCAGCTTGA 1939
QY 533 tcagcaccgattcgctgcgaccgcccacaggtcttctccagagccacgcgcagg 592
DB 1940 TGTTCGGCGTCCCGCCGAGGCGGTGACCCCTGTATGCCCGGGCGGCCAAGACCATCA 1999
QY 593 agagccgcatctgctcgaagacctgtctctccgcacccccacacctggtgccaacccactc 652
DB 2000 ACTTCGGGGTC---CTACGGGATGTCCGCCACCCCTCTCCAGAGCTTCCCATCC 2056
QY 653 tggagaccgagtggttccacggctcgcgcgcaagtggagggcccaacttacaccgcgcg 712
DB 2057 CCTACGAGGAGGCGGTGGCTTCAATTAGCGCTACTTCCAGAGCTTCCCAAGGTGCGGG 2116
713 gccccaatcagggcccgggccttgccacagctgagcgcaagagcggtcgcgcg 772
2117 CCTGGATTGAGAGACCTTGAGGAGGAGGAGGAGGCGGGGTACGTGGAGACCTCTTCG 2176
QY 773 gggacaagagccactt 788
DB 2177 GCCCGCGCGCTACGT 2192

RESULT 15

US-08-756-386-7
Sequence 7, Application US/08756386
Patent No. 5985557

GENERAL INFORMATION:
APPLICANT: Prudent, James R.
APPLICANT: Hall, Jeff G.
APPLICANT: Lyamichev, Victor I.
TITLE OF INVENTION: Invasive Cleavage Of Nucleic Acids
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESS: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America

ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/756,386
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-JUL-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02564
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2502 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-756-386-7

Query Match 1.2%; Score 44.8; DB 2; Length 2502;
Best Local Similarity 46.8%; Pred. No. 0.038;
Matches 176; Conservative 0; Mismatches 197; Indels 3; Gaps 1;

QY 413 tgatgctgcagagcaataagtcgaggagcagaacttgcaggacgacctggtattgtacc 472
DB 1820 TGGTGGCCCTGGACTATAGCCAGATAGAGTCCGGGTCTTGGCCACCTCTCCGGGAGC 1879
QY 473 aggcgtggtgtgagccttcttgaggcgcagccagcatctccgggcgcctacacct 532
DB 1880 AGAACCTGTATCCGGGTCTTCCAGGAGGGAGGACATCCACACCCAGCCAGCTTGA 1939
QY 533 tcagcaccgattcgctgcgaccgcccacaggtcttctccagggccacgcgcagg 592
DB 1940 TGTTCGGCGTCCCGCCGAGGCGGTGACCCCTGTATGCCCGGGCGGCCAAGACCATCA 1999
QY 593 agagccgcatctgctcgaagacctgtctctccgcacccccacacctggtgccaacccactc 652
DB 2000 ACTTCGGGGTC---CTACGGGATGTCCGCCACCCCTCTCCAGAGCTTCCCATCC 2056
QY 653 tggagaccgagtggttccacggctcgcgcgcaagtggagggcccaacttacaccgcgcg 712
DB 2057 CCTACGAGGAGGCGGTGGCTTCAATTAGCGCTACTTCCAGAGCTTCCCAAGGTGCGGG 2116
QY 713 gccccaatcagggcccgggccttgccacagctgagcgcaagagcggtcgcgcg 772
DB 2117 CCTGGATTGAGAGACCTTGAGGAGGAGGAGGAGGCGGGGTACGTGGAGACCTCTTCG 2176
QY 773 gggacaagagccactt 788
DB 2177 GCCCGCGCGCTACGT 2192

Search completed: November 1, 2001, 10:03:06
Job time: 8214 sec

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	380.8	10.4	508	A0764784	HS_3118_A
2	372.4	10.2	534	A0837108	HS_5464_A
3	336.2	9.2	543	A0728799	HS_5464_B
4	222.6	6.1	681	A0717941	AU171941
5	219.2	6.0	280	BB354926	BB354926
6	194	5.3	288	BB429778	BB429778
7	172.6	4.7	895	CNS032UA	AL225307 Tetraodon
8	166.8	4.6	632	CNS02NOT	AL205670 Tetraodon
9	165	4.5	961	CNS022LV	AL178348 Tetraodon
10	108.2	3.0	592	A0881884	AQ881884 HS_5304_A
11	101.2	2.8	877	CNS03H19	AL243702 Tetraodon
12	95.4	2.6	329	A2628814	A2628814 LM0481D17
13	93.2	2.6	1030	CNS03EM8	AL240569 Tetraodon
14	87.2	2.4	988	CNS04BT9	AL283590 Tetraodon
15	86	2.4	633	CNS04FKJ	AL288460 Tetraodon
16	86	2.4	921	CNS03PB5	AL254426 Tetraodon
17	82	2.2	888	CNS03VL7	AL266452 Tetraodon
18	80.8	2.2	807	CNS0317B	AL223184 Tetraodon
19	80.6	2.2	957	CNS020UA	AL207163 Tetraodon
20	80.4	2.2	822	CNS02FTA	AL195463 Tetraodon
21	78.2	2.1	899	CNS032UB	AL225308 Tetraodon
22	77	2.1	483	A2903235	A2903235 RPCT-24-1
23	77	2.1	1096	CNS05CMB	AL331292 Tetraodon
24	75.6	2.1	379	CNS03EM7	AL240568 Tetraodon
25	73.2	2.0	1012	CNS03Y7G	AL265957 Tetraodon
26	72.8	2.0	446	A2870959	A2870959 2M0183F08
27	63.6	1.7	817	CNS05ATS	AL328969 Tetraodon
28	63.4	1.7	925	CNS0091P	AL053013 Drosophila
29	61.8	1.7	591	A0764779	HS_3118_A
30	61.6	1.7	1043	CNS020JP	AL175678 Tetraodon
31	61	1.7	902	CNS006QP	AL065804 Drosophila
32	60.4	1.7	923	CNS0091P	AL053013 Drosophila
33	58.4	1.6	518	A2823042	A2823042 2M0096N15
34	58.2	1.6	1101	CNS012S8	AL101954 Drosophila
35	57.4	1.6	839	CNS004NB	AL054280 Drosophila
36	56	1.5	932	CNS0072Q	AL066742 Drosophila
37	55	1.5	844	CNS0052P	AL056652 Drosophila
38	55	1.5	972	CNS01XPU	AL172011 Tetraodon
39	54.4	1.5	1101	CNS00397	AL063912 Drosophila
40	53	1.5	1100	CNS016KD	AL106855 Drosophila
41	51.6	1.4	1101	CNS017SY	AL108460 Drosophila
42	51.2	1.4	608	BF688560	BF688560 602184874
43	51.2	1.4	918	AQ895329	AQ895329 HS_4832_A
44	51	1.4	839	CNS004NB	AL054280 Drosophila
45	50.6	1.4	817	BE558893	BE558893 HV_CFB002

ALIGNMENTS

RESULT 1	AQ764784	508 bp	DNA	GSS	28-JUL-1999
LOCUS	HS_3118_AL_C10_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-3118 Col-19 Row=E, DNA sequence.				
DEFINITION	AQ764784				
ACCESSION	AQ764784.1	GI:5642900			
KEYWORDS	GSS.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 508) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.				
TITLE	Sequence-tagged connectors: A sequence approach to mapping and				

scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3118 row: E column: 19
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 508.

FEATURES

source
1..508
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3118 Col=19 Row=E"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/notes="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 166 a 129 c 109 g 98 t 6 others
ORIGIN

Query Match 10.4%; Score 380.8; DB 233; Length 508;
Best Local Similarity 96.4%; Pred. No. 9.5e-95;
Matches 407; Conservative 0; Mismatches 13; Indels 2; Gaps 2;

QY 2143 cgggacgagctgaaacactctatgcccaactggaatataaaagaaagagatgc 2202
|||||
DB 88 CAGGACGAGCTGAAAAAATCTATGCCCACTGGAATATATAAAGAAAGATGATC 147
QY 2203 acaaacacccccaccctccagaaaaagcgtgctcgaagaggcctcagtgctccatc 2262
|||||
DB 148 ACAAAACACCCCACTCCACAAAAGCGGTGCTCGAAGAGAGCGCTTANGTCTCCATC 207
QY 2263 atagacgacattacgagatcccgagacacagtcagcgcagtgctctaaagagacaag 2322
|||||
DB 208 ATGAGCGCATTCGAGAGATCCCGAGACAGTACGCGGCGAGTCTCTAAAGAGGACAAG 267
QY 2323 gagggcccgaccatggcagacgcaagggcactgccctcactcaggaagacccccagag 2382
|||||
DB 268 GAGGCGCGGACCATGCGACAGCCAAAGGCACTGCCCTCATCAGAGAACTCCAGAG 327
QY 2383 tcttcagggaacacaggggaaatccaaaggaggagaccctgaaaaaccagtgcttctcactc 2442
|||||
DB 323 TCTTTCATGGAACACATGGAATCCAAAGGAGAGACCTTG-ATAACCGAGTCTTCTCACTC 386
QY 2443 aagaatccccacacacttatgaccacgtgagagacacaaacgagagtcagtagccta 2502
|||||
DB 387 AAGAAATCCACACACACTTATGACACGTGAGAGACCANAGAGAGTCAGTAGCCTA 446
QY 2503 cccacagaaagccagaggagagagacacagaaatccacactggaatccctcgcggt 2562
|||||
DB 447 CCCACAGAAAGCCATGAGGAGAGACACACAG-AAATTNCACACTGGAATCCTCNGGT 505
QY 2563 aa 2564
||
DB 506 AA 507

RESULT 2
AQ837108
LOCUS
DEFINITION
HS_5464.AL_H02.T7A RPCT-11 Human Male BAC Library Homo sapiens genomic clone Plate-1040 Col-3 Row-O, DNA sequence.
ACCESSION
AQ837108
VERSION
AQ837108.1 GI:5806982

KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 554)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1040 row: 0 column: 3
Seq primer: T7
Class: BAC ends
High quality sequence stop: 554.
Location/Qualifiers
1..554
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1040 Col=3 Row=0"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT 182 a 147 c 119 g 102 t 4 others
ORIGIN
Query Match 10.2%; Score 372.4; DB 234; Length 554;
Best Local Similarity 92.0%; Pred. No. 2.2e-92;
Matches 414; Conservative 0; Mismatches 33; Indels 3; Gaps 2;
QY 2157 aaactctatgcccaactggaataataaaagaagaagatgatcacaaaccccca 2216
DB 93 AAAAACTATGCCCACTGGAATATATAAAGAAGAGATGATCACAACACCCCA 152
QY 2217 cctcagaaaaagcgtgtcgaagaaggcctcattcattcattcattcattcattc 2276
DB 153 CCTCAGAAAAAGCGTGTCTGAAGAAAGCGCTANGTCCTCATCATGAGCGATTAC 212
QY 2277 ggaatccagagacagtcagcggcagtcgtcttaaaagagacaaggagggcgcaaca 2336
DB 213 GGAGATCCAGAGACGACGCGCGCAGTGTCTCTAAAGAACAAGAGGGCGCGACCA 272
QY 2337 tggcagacccaaggcactgcctcatcaggaagacccccagagcttcagggaacac 2396
DB 273 TGGCAGACCAAGCACTGCCTCATATGAGAACCCCCCAGAGTCTTCAGGAACAC 332
QY 2397 agggaaatccaggaggagaccctgaaacacgagtcgtcttcactcaagaatccacag 2456
DB 333 AGGGAATTCAGGAGAGACCGCTGAAAAACCGAGTCTTCTCACTCAAGAAATCCACAG 392
QY 2457 cacttataccacgtgagagaccacgaagagtcagtagctaccacagaagcca 2516
DB 393 CACTTATGACCACTGAGAGACCAACCCAGAGTTTACCTTACCCACAGAATGTC- 451

QY 2517 agagaggagacacacagaaattccacactggaatccctgctgggtataaaataacaca 2576
DB 452 AGAAGAGGAGACTACAGAAATCCACACTGGAATTCCTGTCTGGTAAACCTTA--CCC 509
QY 2577 aaaaataaagaagacagcagcgagctgagtc 2606
DB 510 GCTATCTATAGAAGACAGCGGCTTATTC 539
RESULT 3
QY 257799 A0728799 543 bp DNA GSS 15-JUL-1999
LOCUS HS_5464_B2_A11_17A_RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=1040 Col=22 Row=B, DNA sequence.
ACCESSION A0728799
VERSION A0728799.1 GI:5500351
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 543)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1040 row: B column: 22
Seq primer: T7
Class: BAC ends
High quality sequence stop: 543.
Location/Qualifiers
1..543
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1040 Col=22 Row=B"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT 168 a 139 c 124 g 107 t 5 others
ORIGIN
Query Match 9.2%; Score 336.2; DB 232; Length 543;
Best Local Similarity 87.4%; Pred. No. 2.7e-82;
Matches 368; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 2146 gacgagctgaaaaactctatgcccaactggaatataataaaagaagaatgatcac 2205
DB 110 GACGAGCTGAAAAACTCTATGCCCACTGATGTTTCAAGAAAGAGATGATCTCA 169
QY 2206 aaaaacccccaccctccagaaaaaagcgtgctcgaagaaggccttagtctgtccatcatg 2265
|||||

Email: genome-res@rtc.riken.go.jp,
URL: http://genome.rtc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermotabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES

source

Location/Qualifiers
1. .280
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C030003G02"
/clone_lib="RIKEN full-length enriched, adult male corpus
striatum"
/sex="male"
/tissue_type="corpus striatum"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 185.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGATTCGAGTTAATTAATATCCCTCCCTCCCTCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pUC19 with a modified MCS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end:
BamHI"

BASE COUNT
79 a 65 c 62 g 74 t

Query Match 6.0%; Score 219.2; DB 132; Length 280;
Best Local Similarity 86.4%; Pred. No. 8.9e-50;
Matches 242; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 1897 gtcttgctgaagacttcagctgtgagtgatgctgtatttgcacatactcat 1956
Db 1 GTGCTAGCTTCGAGATTCAGCCGATCGATCCCTGCTACTTTGCACATGCTTAC 60
QY 1957 ttgactgtgacagtcaccattgggtgttttggattccaaagtgtttcattcaagaat 2016
Db 61 TTGACTGTGACAGTTACCAGCTGGGTGCTTCTTAATTCAAAGTTTTCATTCAGCAAT 120
QY 2017 aaccacagatgatctgtacagaagcatatgaggatgagctagacatggccgactct 2076
Db 121 RATCCCCGAGAGACATTCCTACGGAACATATGAGATGAGCTGGACATGTTGCTCT 180
QY 2077 ggatcctacctgaacagcagtatcaattcagcctggagtgagcacagcttgatccagag 2136
Db 181 GGATCCTACCTGAACAGCAGTATCAATTCAGCCTGGAGTGAACAGCCTAGATCCAGAA 240
QY 2137 gacattcggagcagctgaagaaactctatgcccactgg 2176
|||||

Db 241 GACATTCGGGTGAAGTGAAGAACTCTATGCTCAGTTAG 280

RESULT 6

BB429778
LOCUS BB429778 288 bp mRNA EST 18-JUL-2000
DEFINITION RIKEN full-length enriched, adult male hippocampus Mus
musculus cDNA clone C630030A14 3', mRNA sequence.
BB429778
ACCESSION BB429778.1 GI:92695505
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 288)
AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata
Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya
T., Tsunoda, Y., Watanabe, S., Watanabe, S., Yamamura, T., Yamanaka, I.,
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino
M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL: http://genome.rtc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermotabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES

source

Location/Qualifiers
1. .288
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C630030A14"
/clone_lib="RIKEN full-length enriched, adult male
hippocampus"
/sex="male"
/tissue_type="hippocampus"
/dev_stage="adult"
/lab_host="SOLR"
/note="Site_1: XhoI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was

[illegible]

	Query Match	2.4%	Score 86;	DB 221;	Length 633;
	Best Local Similarity	73.3%	Pred. No. 1.7e-12;		
	Matches 110;	Conservative	0;	Mismatches 40;	Indels 0; Gaps 0;
QY	1997	agttttcacattcaagcaataaccacagagatgattgtctacagagcatatggaatg	2056		
Db	343	AGTTCTCTTTGGCGGGACCCACATGTGGAGACGACATGGCCCTCGAGGCCCTACGAGACG	402		
QY	2057	agcttagacatgggcccgatctggatcctcactgaacagcagtcagtaaatcgacctcggagtg	2116		
Db	403	AGCTGGACATGGGGCGCTCCGGGTCTACCTCAACACGACGATCACGTCGGCGTGGAGCG	462		
QY	2117	agcacagcttggattccagagagacaattcggg	2146		
Db	463	AGCACGCTTGGATCCCGGAACACATTCGGG	492		

Search completed: November 1, 2001, 09:12:23
Job time: 8831 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 09:12:27 ; Search time 45.13 Seconds
(without alignments)
1632.133 Million cell updates/sec

Title: US-09-775-181-2

Perfect score: 6382

Sequence: 1 MGNAYPPLLCLLLAQLGLG.....LSANKIAPRKEEIWDSFKV 1215

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601.*
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	223.5	3.5	1199	13 AAR25080	GLU-G-R subtype 1a
2	220	3.4	1219	18 AAW25763	Amino acid sequenc
3	212	3.3	1197	20 AAV28562	Mouse metabotropic
4	210	3.3	1171	20 AAV28563	Rat metabotropic g
5	198	3.1	1180	16 AAR64253	Human mGluR5a. Ho
6	198	3.1	1397	20 AAV49134	pmGluR2/Car*Galpa
7	197.5	3.1	1212	20 AAV28564	Human metabotropic
8	193.5	3.0	1212	16 AAR64254	Human mGluR5b. Ho
9	193.5	3.0	1394	20 AAV49129	pmGluR2/Car*Galpa
10	193	3.0	1194	14 AAR42199	HsmGluR1. Homo sa
11	190	3.0	1056	18 AAW25765	Amino acid sequenc

12	186	2.9	1056	18 AAW25764	Amino acid sequenc
13	183.5	2.9	1276	20 AAY49127	phCar/hmGluR2*Gq15
14	180	2.8	906	13 AAR25081	GLU-G-R subtype 1b
15	179.5	2.8	3266	21 AAB42491	Human ORFX ORF2255
16	178	2.8	571	19 AAW55064	Streptococcus pneu
17	177	2.8	906	16 AAB64250	Human mGluR1B. Ho
18	176	2.8	917	20 AAY49126	Chimeric hCAR/hmG1
19	176	2.8	1035	20 AAY49138	Human myosin I-cua
20	172.5	2.7	982	19 AAW1632	Rat kidney calciu
21	172.5	2.7	1079	19 AAW54847	Rat kidney cell ca
22	172.5	2.7	1079	19 AAW38275	Rat parathyroid ca
23	172.5	2.7	1079	20 AAY41781	Rat parathyroid ca
24	172.5	2.7	1079	20 AAW94928	Rat kidney extrace
25	172.5	2.7	1079	20 AAW89566	Rat parathyroid ca
26	172.5	2.7	1079	21 AAY51828	Rat calcium recept
27	172.5	2.7	1418	20 AAY49131	mGluR8/Car*Galpha
28	170	2.7	877	16 AAR64255	Human mGluR5c. Ho
29	169	2.6	554	19 AAY20763	Human neurofilamen
30	169	2.6	974	21 AAY45000	Human calcium sens
31	168.5	2.6	1257	22 AAB66926	RBPL protein. Hom
32	168	2.6	1863	21 AAY77818	BRCA1 protein sequ
33	167.5	2.6	682	21 AAG46506	Arabidopsis thalia
34	167.5	2.6	724	21 AAG46505	Arabidopsis thalia
35	167.5	2.6	778	21 AAG46504	Arabidopsis thalia
36	167	2.6	1863	18 AAW23286	Human breast and o
37	166	2.6	1114	21 AAB08517	Amino acid sequenc
38	166	2.6	1596	18 AAW31347	Rat tumour suppres
39	164	2.6	1965	19 AAW69165	Streptococcus pneu
40	163.5	2.6	480	21 AAY82317	Human protein tran
41	163.5	2.6	877	16 AAR82657	Human mGluR3. Hom
42	163.5	2.6	901	21 AAY45001	Human calcium sens
43	163	2.6	718	21 AAG42569	Arabidopsis thalia
44	163	2.6	738	21 AAG42568	Arabidopsis thalia
45	163	2.6	763	21 AAG42567	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAR25080
ID AAR25080 standard; Protein; 1199 AA.
XX AAR25080;
XX
XX
DT 04-JAN-1993 (first entry)
XX
DE GLU-G-R subtype 1a.
XX
KW Xenopus oocyte; cerebellum sublibrary; voltage clamp assay;
KW neurotransmitter; glutamate; neuronal excitation; clone 45-A.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 98 /note= "N-glycosylation"
FT Modified-site 223 /note= "N-glycosylation"
FT Modified-site 397 /note= "N-glycosylation"
FT Modified-site 515 /note= "N-glycosylation"
FT Domain 594..614 /label= Transmembrane_domain
FT Domain 630..650 /label= Transmembrane_domain
FT Domain 657..677 /label= Transmembrane_domain
FT Domain 707..727 /label= Transmembrane_domain
FT Modified-site 747 /note= "N-glycosylation"
FT

FT Domain 753..772
FT /label= Transmembrane_domain
FT 787..807
FT /label= Transmembrane_domain
FT 812..832
FT /label= Transmembrane_domain
FT 920
FT Modified-site /note= "N-glycosylation"
FT 925
FT Modified-site /note= "N-glycosylation"
FT 1180
FT Modified-site /note= "N-glycosylation"
FT
XX W09210583-A.
XX
XX
XX 25-JUN-1992.
XX
XX 12-DEC-1991; 91WO-US09422.
XX
XX 12-DEC-1990; 90US-0626806.
XX 30-JAN-1991; 91US-0648481.
XX 18-MAR-1991; 91US-0672007.
XX
XX (UNIW) UNIV WASHINGTON.
XX (ZYMO) ZYMOGENETICS INC.
XX
XX
XX Almers W, Hagen FS, Houamed KM, Mulvihill ER;
XX
XX WPI; 1992-234638/28.
XX N-PSDB; AAQ25812.
XX

Mammalian G protein-coupled glutamate receptors - activate phospholipase C or stimulate inositol phospholipid metabolism, for use in diagnosis and identification of receptor agonists and antagonists

Disclosure; Fig 5; 144pp; English.

The sequence given is mammalian G protien-coupled glutamate receptor (GLU-G-R) subtype 1a. The clone encoding this protein was isolated from Xenopus oocyte cultures which had been injected with RNA extracted from a cerebellum sublibrary. Expression of GLU-G-R activity was assessed by voltage clamp assay. Plasmid DNA encoding GLU-G-R activity was isolated by replica-plating experiments. The protein encoded by clone 45-A was found to act as a receptor for the neurotransmitter glutamate, and may play a critical role in the control of neurons. GLU-G-R and antibodies may be used to identify agonists and antagonists of GLU-G-R-mediated neuronal excitation and in diagnosis.

Sequence 1199 AA;

Query Match 3.5%; Score 223.5; DB 13; Length 1199;
Best Local Similarity 18.6%; Pred. No. 1.6e-07;
Matches 164; Conservative 121; Mismatches 325; Indels 271; Gaps 28;
QY 357 GFYHVGVLVNNFRRRGPDQHISGSTKDVSEE-----AYVCLPCREG- 398
DB 498 gtwhgevlnddykqkq--mnksgmvrsvscpcqkqikvirkgevsccwictackene 554
QY 399 -----CFPC-----ADDSCFVQEDKYLK-----LAIISFGCLMLDFVSMVLV 438
DB 555 fvqdeftracdlgwwpnaeltgceipvrylewsadlesilalafscqilvltlflif 614
QY 439 YHFRKAKSIRASGLILETLFGSLLLYPPVVLVFEPTFFRCILLRWARLLGFATVYGT 498
DB 615 vlyrdtpvkkssrelcyililagiflygvpcftllakpttscyqrlivglissamcysa 674
QY 499 VTLKLRVLKV-----FLSRTAQRIPTYMTGGRYMRMLAVILLVFWFLIGW 544
DB 675 lvtktnrariilagskkkictckprfmsawaqvaiilaisvqltlvltlii----- 726
QY 545 TSSVCONLEKQISLGQGTSTDHLLFNMCILDRDYMATAVA-EFLFLMGVYLCYAVRTV 603

DB 727 -----meppmpillspsike--vylcntsnlgvavpgvngllimscyyafktrnv 777
QY 604 PSAPHEPRYMAVAVHNELIISAIFHTIRFVLASRLQSDMLMLYFAHTHLTVTVTIGLL 663
DB 778 panfneakyiafmyttclwafvpiyf-----gsnykllttcfavslsvcvaigcmf 831
QY 664 IPK----FSHSSNNPRDDIATEAYEDELDMG-----RSGSYLNSNSINSAWSEHSLDP 711
DB 832 tpkmyliiakpernvrsaft--sdvvrhvgdgklpcrsntfln----- 874
QY 712 EDIRDELKLYAQLEIYKRKKMITNPNHLOKRC-----KKGIGRIMRRITIPET 764
DB 875 -----ifrrtkpgagnansgksvswpggqgkqghvwqrlsvhvkt 919
QY 765 VSRQCKEDKEGADHGTAKGTALIRKNPPESSGNTCKSEETLKNRVFSLKKSHTYDHY 824
DB 320 netaenq-----tavik-----pltksy----- 937
QY 825 RQOTESSSLPTESQDEETTENSTLESGLKLTQKLKEDSEASTESVPLVCKKSASAHN 884
DB 938 -----qsgnsltfsdastklynvveedntpsahfppspmvvhr 980
QY 885 LSSEKKTGHPRTSMLOKLSVIAAKEKTLGLAGKTQTAGVEERTKSKQKPLPKDKETNRN 944
DB 981 rgppvattppplphl-----taeetplfadsvipkgl-----ppplpqgq----- 1021
QY 945 HNSDNTETKDPAPQNSNPAEPRKPKQSGIMKQQRVNTTANSDL-----NPG 993
DB 1022 -----pqqpqpqpqpqpqk--slmdqlgvvtnfgsgipdfhavlagnptg 1066
QY 994 TTQMKDNFDIGEVCPWEVYDLTPGPVPSE--SKVQKHVSIVASEMEKNTFSLKKESSHK 1051
DB 1067 nt-----lrslypppppqhlqmplhlstfgeesisppgediddds--- 1108
QY 1052 PRAAEVCQOSNOKRIDKAEVCLWESQG-----QSILEDKLLISKTPVLPRAKEENGQQR 1108
DB 1109 -----erfkllqefvyeregnteedeleeedlptaskitpedspaltppsp 1156
QY 1109 AANVCAGQSEELPP--KAVASKTENENLQIGHOKTSSS 1147
DB 1157 rdsvasgssvpspsvsvlctppnvtyasvillrdykqsss 1197
RESULT 2
AAW25763
ID 'AAW25763 standard; Protein; 1219 AA.
XX AAW25763;
XX
XX 03-DEC-1997 (first entry)
XX Amino acid sequence of pCar/R1.
XX
KW Chimeric receptor; extracellular domain; seven transmembrane domain;
KW intracellular cytoplasmic tail domain; metabotropic glutamate receptor;
KW MGR; calcium receptor; CR; mGluR; G protein-coupled receptor;
KW neurological disease.
XX
OS Chimeric - Homo sapiens.
XX
XX WO9705252-A2.
XX
XX 13-FEB-1997.
XX
XX 25-JUL-1996; 96WO-US12336.
XX
XX 26-JUL-1995; 95US-0001526.
XX
XX (NPSF-) NPS PHARM INC.
XX
XX Fuller FH, Hammerland LG, Krapcho KJ;

Best Local Similarity 18.5%; Pred. No. 1.1e-06;		Matches 177; Conservative 126; Mismatches 402; Indels 252; Gaps 35;	
QY	219	ETWFHGLRRKWRPHLHRRGNQGRGLGHSWR----	RKDGGLGDKSHFKWSPPYLECE 273
Db	332	dvkwdyylklrpetnlrnp-----wfgfwhrfqcrlegfaqenskyntk-----cn	381
QY	274	NG-----SYKPGWLVTLSAIYGL---QPNLVPEFRGV-----MKVDINLQ	311
Db	382	ssltlrthhvqskmgfvinaismayghnmqmslcpagyagldcmkpidgrkllidslm	441
QY	312	KVDIDQCSDDGWFSGTHKCHLNNSCMPKGLGVGLGAYECICKAGFYHPGVLPV-----	366
Db	442	ktnftvgsgdmllfdengdsparyelmnfkemg---kdyfdivngswdngelkmddev	498
QY	367	-----NNFRRRGPDOHI-SGSTKDVSEE-----AYVCLPCREG-----CPFC	403
Db	499	wsknnlirsvcspecekgqkvrkgevsccwtctpckeneyvfydeyckacqlgsupt	558
QY	404	DD-SPCFVQEDKYLR-----LAIISFQGLCMLLDFVSMVYVYHFRKAKSIRASGLILLE	456
Db	559	ddltgdcldlpvqylrwgdppeiaavfacglilatitfvtfviifdyrdtpvksrrelcy	618
QY	457	TILFGSLLLYFPVILYFPEPSTFRCLLRWALLGFATVGTWTLKLRVLKV-----	509
Db	619	illaglcylctfcilakpkqicylqrlgiglspsamsyalvtktnriarilagskkk	678
QY	510	-----FLSRTAQRIPYTGGRVWRMLAVILLVWPFILGWTSSVCNLEKQISL----	558
Db	679	icckkprfmsacq-----lviafili-----ciglglilalfime	714
QY	559	---IGOGKTSDDLIFNMCLDRWDYMTAVA-EPLFLMWGYLYCYAVRTVPSAFHEPRYMA	614
Db	715	ppdimbdpsirevylcnctnlgtvptlgyngllilscftvafktrnpanpanfneakyla	774
QY	615	VAVNELIISAIPHTTRFVLASRLQSDMMLMYFAITHUTVTVTIGLLIPK-----FSHS	670
Db	775	fmytctiilawfviyf-----gsnykiltmcfsvsatsatvalgcmfvkpyvillakp	828
QY	671	SNPRDIIATE-AYEDELDMGRSGYL--NSSINSAMSEHSLDPEDIRDELKKL----	724
Db	829	ernvrsaftstvmnhvgdgksssaarssslvnlwkrsgsgetlrykdrllaqhsk	888
QY	725	LEIYKRKMITNPHLQKRCSSKGL-----GRSIRRR-----I	758
Db	889	iecfpkpgmgnggratmssngskvtwaqnekstrgqhlwqlslvhinkkenpqtavi	948
QY	759	TETPETSVCCKEDKEGADHGTAQ-----TALIRKNPPSSSGWTGKSKEETLKN	809
Db	949	kfpkate---arggagagaggpgaaagagactatggpeppdagka-----	995
QY	810	RVFLSKKSHSTVDHVRDQTEESSSLPTESQEEETTENTLESLSGKKLTQKLKEDSEAS	869
Db	996	-----lydvaeeerfpaarprspisltishlag-----sagrt	1031
QY	870	TESVPLVCKSASAHNLSSKKTGHPRTSMLOKSLSVSIASAKEITLGLAG-----	918
Db	1032	dddaalshtaarssssgslmeqissvvtftaniteinsmmistaaapppgtpics	1091
QY	919	-----KTQTAGVETRKQKPLPKDKETNRNHSNSENTETKDAPQNSNPAEP	967
Db	1092	sylikelqlptmtttfae-----iqplalevt-----ggaqpatgspaqetpagae	1140
QY	968	RKPKSGIMKQQRVNPPTTANSLNPOTQMKDNFDIGEVCPWEVVDLTPGPVPSESK	1024
Db	1141	aapgkpdleelvaltp-----psfrsvdsgsttpuspsesalcipsspk	1187

RESULT 4
 ID AAY28563
 XX AAY28563 standard; Protein; 1171 AA.
 AC AAY28563;

XX	09-NOV-1999	(first entry)
XX		
DE		Rat metabotropic glutamate receptor 5 (mGluR5).
XX		
KW		Metabotropic glutamate receptor; mGluR5; central nervous system; CNS;
KW		depressant; anaesthesia; phospholipase C; intracellular calcium.
XX		
OS		Rattus sp.
XX		
PN	W0938975-A2.	
XX		
PD	05-AUG-1999.	
XX		
PF	29-JAN-1999;	99WO-US02033.
XX		
PR	30-JAN-1998;	98US-0073190.
XX		
PA	(UYTE-) UNIV TECHNOLOGY CORP.	
XX		
PI	Johnson TE, Rikke BA, Sikela JM, Simpson VJ;	
XX		
DR	WPI: 1999-494093/41.	
DR	N-PSDB; AAZ10803.	
XX		
PT	CNS depressant sensitivity associated a YAC containing murine	
PT	tyrosinase and metabotropic glutamate receptor 5 genes, useful for	
PT	identifying CNS depressant modulators	
XX		
PS	Claim 35; Fig 16; 140pp; English.	
XX		
CC	This sequence is the rat metabotropic glutamate receptor 5 (mGluR5).	
CC	This sequence is not directly claimed, but a claimed polypeptide (part	
CC	of AAY28562:mouse mGluR5) is specifically not depicted in the rat	
CC	mGluR5. mGluR5 are G protein-coupled receptors capable of activating a	
CC	variety of second messenger systems. mGluR5 belongs to mGluR Group I. The	
CC	binding of an agonist to group I receptors results in the activation of	
CC	phospholipase C causing the mobilisation of intracellular calcium. The	
CC	mouse mGluR5 sequence correspond to parts of the mouse vrt2	
CC	polynucleotide which is associated with central nervous system (CNS)	
CC	depressant activity. CNS depressants are used for a variety of purposes	
CC	including anaesthesia.	
XX	Sequence	1171 AA;
SQ		
Quéry Match 3.3%; Score 210; DB 20; Length 1171;		
Best Local Similarity 19.1%; Pred. No. 1.5e-06;		
Matches 179; Conservative 117; Mismatches 397; Indels 244; Gaps 35;		
QY	219	ETWFHGLRRKWRPHLHRRGNQGRGLGHSWR-----RKDGGLGDKSHFKWSPPYLECE 273
Db	332	dvkwdyylklrpetnlrnp-----wfgfwhrfqcrlegfaqenskyntk-----cn 381
QY	274	NG-----SYKPGWLVTLSAIYGL---QPNLVPEFRGV-----MKVDINLQ 311
Db	382	ssltlrthhvqskmgfvinaismayghnmqmslcpagyagldcmkpidgrkllidslm 441
QY	312	KVDIDQCSDDGWFSGTHKCHLNNSCMPKGLGVGLGAYECICKAGFYHPGVLPV----- 366
Db	442	ktnftvgsgdmllfdengdsparyelmnfkemg---kdyfdivngswdngelkmddev 498
QY	367	-----NNFRRRGPDOHI-SGSTKDVSEE-----AYVCLPCREG-----CPFC-----A 403
Db	499	wsknnlirsvcspecekgqkvrkgevsccwtctpckeneyvfydeyckacqlgsupt 558
QY	404	DD-SPCFVQEDKYLR-----LAIISFQGLCMLLDFVSMVYVYHFRKAKSIRASGLILLE 456
Db	559	ddltgdcldlpvqylrwgdppeiaavfacglilatitfvtfviifdyrdtpvksrrelcy 618
QY	457	TILFGSLLLYFPVILYFPEPSTFRCLLRWALLGFATVGTWTLKLRVLKV----- 509
Db	619	illaglcylctfcilakpkqicylqrlgiglspsamsyalvtktnriarilagskkk 678

```
QY 510 -----FLSRTAQRIPYMTGGRRVMMRLAVILLVWFVFLIGWTSVCONLEKQISL----- 558
DB 679 ictkpkfrmsacq-----lviafill-----ciqgiilvalfime 714
QY 559 ---IGOGKTSDDLIFNMCLIDRWDMYTAFA-EFLFLLGWVYLCYAVRTVPSAPHEPYMA 614
DB 715 ppdmdhypsirevylicntnlgvvplgnyngllilscfyafrktrnvpantneakyla 774
QY 615 VAVHNELIISAIFHTIRFVLASRLQSDWMLMLYFAHPLHVTVTIGLLLPK-----PSHS 670
DB 775 fmytctiilwafviyf-----gsnykiitmcfsvsatvalgcmfvpkyvillakp 828
QY 671 SNPNRRDIATE-AYEDELDMGRSGSYL--NSSINSAMSEHSLDPEDIRLEKK----- 720
DB 829 ernvrsaftstsvrmhvgdgkssaaarssslvnlwkrsgsgetlssngksvtwaqne 888
QY 721 -----LYAQLEIYKRRKMTNNPHLQK---KRCSSKGLGRSMTMRITPEITVSRQCS 770
DB 889 kstrgqhlwrlsvhlnkpenqatavikpfpkstenrg-----paaaagggg 936
QY 771 KEDKEGADHGTAKGTALIRKNPNPSSGNTGKSKEETLKNRVFSLKSHSTYDHRDQTEE 830
DB 937 gpgvaga--gnagctatgpgpepdagpka-----lydvaea 970
QY 831 SSSLPTESQEEETENSTLESGLKLTQKLKEDSEAESESTESVPLVCKSAINLSSEKK 890
DB 971 eesfpaarprspistlshlag-----sagrtddapslhsetaarssssgs 1020
QY 891 TGHPRTSMLQKLSVIASAEKTLGLAG-----KTQTAGVEER 928
DB 1021 lmeqlssvvtfrtaniselnmmlstaatpgppgtpicssyllipkeiqblptmtfae-- 1078
QY 929 TKSQKPLPKDKEYNRHNSNDNTEKDPAPONSNAPEE-PRKPKQSGIMKQQRVNPPTAN 987
DB 1079 ---iqplpaievt-----ggaagatgvsyagdetptgaesagpkpdlleialtppsp--- 1127
QY 988 SDLNPGTQMKNDPFDIGEVCPEWYDITLTPGPVPSSEK 1024
DB 1128 -----frdsvdsgsttpnspvsesaicipsk 1155

RESULT 5
AAR64253
ID AAR64253 standard; Protein; 1180 AA.
XX
AC AAR64253;
XX
21-JUL-1995 (first entry)
Human mGluR5a.
DE
XX
KW Metabotropic glutamate receptor; mGluR5; mGluR5a; hippocampus;
KW cerebellum; PCMV-r7-3.
XX
OS Homo sapiens.
XX
XX WO9429449-A.
XX
XX PD 22-DEC-1994.
XX
XX 03-JUN-1994; 94WO-US06273.
XX
XX 04-JUN-1993; 93US-0072574.
XX
XX (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.
PA
XX Daggett L, Ellis SB, Hess SD, Johnson EC, Liaw C;
XX Pontsler A;
XX
XX WPI; 1995-036478/05.
DR
DR N-PSDB; AAQ80419.
XX
```

```
PT New DNA encoding human metabotropic glutamate receptor
PT sub:types(s) - and related proteins, probes, RNA, transformed
PT cells and antibodies, useful in treatment, diagnosis and
PT identification of specific modulators
XX
XX Claim 15; Page 90-95; 125pp; English.
PS
XX
CC A human hippocampus cDNA library was screened with DNA encoding
CC the rat mGluR1 receptor. An isolated clone was used to screen
CC human cerebellum cDNA libraries to isolate additional clones. 3
CC 3 splice variants of the human mGluR5 transcript, mGluR5a (given
CC in AAQ80419), mGluR5b (AAQ80520) and mGluR5c (AAQ80421), encoding the
CC proteins given in AAQ80520-55, respectively, were constructed by
CC ligating different overlapping clones. Coding sequences are linked
CC to regulatory elements of pCMV-r7-3 for expression in mammalian
CC cells.
XX
SQ Sequence 1180 AA;
```

```
Query Match 3.1%; Score 198; DB 16; Length 1180;
Best Local Similarity 18.9%; Pred. No. 1.2e-05;
Matches 178; Conservative 116; Mismatches 401; Indels 248; Gaps 33;

QY 219 ETWFHGLRRKWRPHLRHRRGNQGRGLGHSRRKDGGLGDKSHFKWSPPYLECENSY- 277
DB 333 dvkwfddyiklrpetnhrp-----wfgewqhr-----fqrleafqdenskyn 378
QY 278 -----KPGMLV-TLSSAIYGL---QPNLVPERGV-----MKVD 307
DB 379 ktcnssltkthvqdkmgfvinaiysmayglhnmqmsicpgyagldcamkpidgrkll 438
QY 308 INLOKVIDIDOCSSDGWFSGTHKCHLNNECMPIKGLGFLVGAYECICAGFYHFGVLPVN 367
DB 439 eslmkntftvgdgtlfdengdsprgyelmmfkemg-----kdyfdyinvgsd 488
QY 368 NFRRRGPDQHI-----SGSTKDVSEE-----AYCULPCREG----- 398
DB 489 ngelkmdddewskksnliirsvcepcekgqikvirkgvscwtctpckeneyvfdet 548
QY 399 CPFC-----ADD-SPCFVQEDKYLK-----LAILSFQGLCMLLDVSMVLYVHFRKAK 445
DB 549 ckacqlgswptdtdlgtcdilipvqylrwgdpapiaavfacilgllptlftvfvfilyrdtp 608
QY 446 SIRASGLILLETILFGSLLLYFPVILYFPSTFRICILLRWALLFPATVGTVTLLKLR 505
DB 609 vvkssrelcyilagicylctfcliaqpkqilycylqriglspamsyalvtktmr 668
QY 506 VLKV-----FLSRTAQRIPYMTGGRRVMMRLAVILLVWFVFLIGWTSVCON 551
DB 669 iarilagskkkictpkprfmsacq-----lviafill-----ciq 704
QY 552 LERQISL-----IGOGKTSDDLIFNMCLIDRWDMYTAFA-EFLFLLGWVYLCYAVRTV 603
DB 705 lgiivalfimeppdmdhypsirevylicntnlgvvptpgnngllilscfyafrktrn 764
QY 604 PSAPHEPYMAVAVHNELIISAIFHTIRFVLASRLQSDWMLMLYFAHPLHVTVTIGLL 663
DB 765 panfpeakylaftmytctiilwafviyf-----gsnykiitmcfsvsatvalgcmf 818
QY 664 IPK-----FSHSSNNPRDIATE-AYEDELDMGRSGSYL--NSSINSAMSEHSLDPEDIRD 716
DB 819 vpkvyiilakpernvrsaftstsvrmhvgdgkssaaarssslvnlwkrsgsgetlss 878
QY 717 ELKK-----LYAQLEIYKRRKMTNNPHLQK---KRCSSKGLGRSMTMRIT 759
DB 759 gngksvtwaqneksrgqhlwgrlshlnkpenqatavikpfpkstesrglag----- 932
QY 760 EIPETVSRQSKEDKADHGTAKGTALIRKNPNPSSGNTGKSKEETLKNRVFSLKSHS 819
DB 933 -----agag-gsagvgatgagcagagpggpespdagptkaly----- 969
QY 820 TYDHRDQTEESSSLPTESQEEETENSTLESGLKLTQKLKEDSEAESESTESVPLVCKS 879
```

Db 970 -----dvaeeahfpaprrpspsistshrag-----sastddvpslhseparvs 1019
 QY 880 ASAHNLSSSEKKTGHPR-TSNLQKSLSVIASAKEKTGLAG-----KTQTA 923
 Db 1020 ssggsimeqissvvtrftaniselnsmlstaapspgvaplcssylipkeiqtlptmtt 1079
 QY 924 GVEERTKSKPLPKDKETNRNHSNNTETKDPAPQN--SNPAEPRKPKQKSGIMKQQRV 981
 Db 1080 fae-----lqplalevtggaqaagaagaadarespaagpeaaaakpdeelevaltpp 1134
 QY 982 NPTTANSLNPGTOMKDNFDJGEVCPWEVYDITPCPVPSSEK 1024
 Db 1135 sp-----frdsvdsgsttpnpsvesalcpsspk 1164

RESULT 6

AA49134
 ID AAY49134 standard; Protein; 1397 AA.

XX AAY49134;

07-JAN-2000 (first entry)

DE pmGluR2/Car*Galphai5+3Ala linker fusion construct protein sequence.

XX G-protein fusion receptor; Car; calcium receptor; GluR; head injury;
 KW metabotropic glutamate receptor; GABABR; chimeric receptor; stroke;
 KW gamma-aminobutyric acid receptor; allosteric modulator; antagonist;
 KW spinal cord injury; epilepsy; ischaemia; hypoglycaemia; anoxia;
 KW Alzheimer's disease; hyperparathyroidism; osteoporosis; depression;
 KW cognitive disorder.

XX Homo sapiens.

XX WO951641-AL.

XX 14-OCT-1999.

XX 02-APR-1999; 99WO-US07333.

XX 03-APR-1998; 98US-0080671.

XX (NPSF-) NPS PHARM INC.

XX Stormann TM, Hammerland LG, Storjohann LL, Busby JG, Garrett JB;
 PI Simin RT;

XX WPI; 1999-610995/52.

XX N-PSDB; AAZ31065.

PT New G-protein fusion receptors and chimeras containing domains from
 PT different receptors, used to screen for modulators, potentially useful
 PT e.g. for treating or preventing stroke or Alzheimer's disease

XX Example 1; Fig 12; 255pp; English.

XX The invention relates to G-protein fusion receptors (I) comprising:
 CC (1) in the N to C direction, extracellular (ECD), transmembrane (TMD) and
 CC intracellular (ICD) domains, each chosen independently from a Car
 CC (calcium receptor), GluR (metabotropic glutamate receptor) and GABABR
 CC (gamma-aminobutyric acid receptor); (2) an optional linker attached to
 CC the C-terminus of ICD; and (3) a G-protein (GP) linked to ICD or the
 CC linker. (I), and recombinant chimeric receptors (CR) without the GP
 CC component, are used to assess function of the various domains and to
 CC identify compounds (e.g. allosteric modulators or antagonists) that act
 CC on these domains. The modulators are potentially useful for treating or
 CC preventing diseases associated with the receptors, e.g. stroke, head or
 CC spinal cord injury, epilepsy, ischaemia, hypoglycaemia, anoxia,
 CC Alzheimer's disease, hyperparathyroidism, osteoporosis, cognitive
 CC disorders and depression. Nucleic acid (II) that encodes (I) is used:
 CC (1) for recombinant production of corresponding proteins; and (2) to
 CC produce cells used in screening for modulators. Use of Car and mGluR

CC domains allows presentation of GABABR domains, to a binding agent, in a
 CC form more like the natural domain structure compared with use of
 CC incomplete receptors, lacking one or more domains. By shuffling different
 CC domains, agents can be identified that affect particular domains of a
 CC receptor.

XX Sequence 1397 AA;

Query Match

Best Local Similarity 3.1%; Score 198; DB 20; Length 1397;

Matches 131; Conservative 88; Mismatches 236; Indels 148; Gaps 27;

QY 390 YVCLPREGCPF-----CADD-----SPCFVQEDKYLK-----LAIFS 422

Db 520 wlicpcq--pyeyrldefcadcglgywnaslfgcfcelpqeyirgdawavgpytiac 576

QY 423 FQGLCMLLDFFVSMVYHFRKAKSTRASGLILLETILFGLSLLLYFPVVTLYPEPSTFRCI 482

Db 577 lgalatl--fvlgvfrh--natpvvkasgrelycillgsvficycmftfiakpstavct 633

QY 483 LLRWARLLGFATVYGVTKLHRLVKVF--LSRTAQRIPYMT--GGVVRMLAVI---LL 535

Db 634 lrrlglgtafsvcsailtktnriariffgaragqrprfispasqvalciaisgqlli 693

QY 536 VVFWFLI---GWTSSVQCNLEKQISLIGQKTSQD---HLIFNCLIDRWDMYMTAVAEFLF 589

Db 694 vvawlvveapgtgketaperrevvtlrcnhrdasmglslaynvlli----- 739

QY 590 LLMGVYLC--YAVRT--VPSAFHEPRYMAVAVHNELIIISAIPTHTFRVLAS--RLQSDMM 643

Db 740 -----alctlyafktrkpcnfneakfigfmyttciliwlaflpfyvtssdyrvgttm 794

QY 644 LMLYFAHTHLTVTVTIGLLLPKF-----SHSSNNPRDIDATEYEDELDM----- 689

Db 795 cv-----svslsgsvvlgclfpaklhilifqpknkieevrcstaaahkvaaratlrnsn 850

QY 690 -----GRSGSLNSSINSAMSEHSLDPEDIRDELKLYAQLEIYVKKKMIT---N 736

Db 851 vskrssslggstgstpsissksnsedpfqperqkqqlaltqceqqqqltlpqq 910

QY 737 NPHLOKKRCSKK---GLGSRIMRRITEIPETVSVROCKEDKEGADHGTA-----K 783

Db 911 qrsqgqprckqkvifsg-----tvtfslsfdepqknamahgnstqhnsleaqs 960

QY 784 GTALTRKNP--PESSGNTGKS---KEETLKNRVFSLKKSHSTYDHDVDRDQTESSSLPTES 838

Db 961 sdtlrrhqlpllcqgetdldltvqetglqgvvggqdr-----pevedpeelspalvvs 1015

QY 839 QBEETTENSTLESLSGKKLTQKLKEDSEASTESVPLVCKSASAHNLSSSEKKTGHPRTSM 898

Db 1016 sq-----sfvisgggstvtenvvnsaaamtlesimacc-----lseeakearinde 1062

QY 899 LOKSL 903

Db 1063 lerql 1067

RESULT 7

AA49134

ID AAY28564 standard; Protein; 1212 AA.

XX AAY28564;

XX 09-NOV-1999 (first entry)

XX Human metabotropic glutamate receptor 5 (mGluR5).

DE Metabotropic glutamate receptor; mGluR5; central nervous system; CNS;

KW depressant; anaesthesia; phospholipase C; intracellular calcium.

XX Homo sapiens.


```

ID AAW25764 standard; Protein; 1056 AA.
XX AC
XX AAW25764;
XX XX
DT 03-DEC-1997 (first entry)
XX DE
XX Amino acid sequence of pratch3.
XX DE
XX Chimeric receptor; extracellular domain; seven transmembrane domain;
KW intracellular cytoplasmic tail domain; metabotropic glutamate receptor;
KW MGR; calcium receptor; CR; mGluR; G protein-coupled receptor;
KW neurological disease.
XX XX
XX Chimeric - Rattus rattus.
XX OS
XX W09705252-A2.
XX PN
XX 13-FEB-1997.
XX PD
XX 25-JUL-1996; 96WO-US12336.
XX XX
XX 26-JUL-1995; 95US-0001526.
XX XX
XX (NPSF-) NPS PHARM INC.
XX PA
XX Fuller FH, Hammerland LG, Krapcho KJ;
XX PI
XX WPI: 1997-145690/13.
XX DR
XX N-PSDB; AAT86167.
XX XX
XX Chimeric receptors comprising metabotropic glutamate receptor and
XX calcium receptor - used for screening for neurologically active
XX compounds
XX XX
XX Example 4; Fig 4; 177pp; English.
XX XX
XX This sequence represents a chimeric receptor. The chimeric receptor
XX comprises an extracellular domain, a seven transmembrane domain, and
XX an intracellular cytoplasmic tail domain, and a sequence of at least
XX 6 contiguous amino acids is homologous to a sequence of a metabotropic
XX glutamate receptor (MGR), and a sequence of at least 6 contiguous amino
XX acids is homologous to a sequence of a calcium receptor (CR). The
XX chimeric receptor may be used for screening for compounds that bind to
XX or modulate the activity of MGR or CR and for determining the site-of-
XX action of a CR active compound. The compounds can be used in the
XX treatment of neurological diseases and disorders. They can also be
XX used as diagnostic agents. Chimeric receptors such as this, allow the
XX coupling of certain functional aspects of an MGR with certain functional
XX aspects of a CR. They allow for more efficient high-throughput screening
XX of compounds.
XX SQ
XX Sequence 1056 AA;

Query Match 2.9%; Score 186; DB 18; Length 1056;
Best Local Similarity 19.7%; Pred. No. 7.7e-05;
Matches 123; Conservative 98; Mismatches 241; Indels 162; Gaps 20

QY 357 GYHFGVLVFNFRRGDPDHIHSGTKDVSSE-----AYVCLPCREG- 398
Db 498 gtwhgegvlnddykikq---mknsgmvrsvscsepclkgqivirkgevsccwictackene 554
QY 399 -----CPFC-----ADSPCFVQEDKYLK-----LAIISFGGLMLLDFVSLVV 438
Db 555 fvqdeftracrdlgwvnaeitgcepiprvylewsdiesiaiafscdglvltlftlfl 614
QY 439 YHFRKAKSIRASGLILLETILFGSULLYFPVVIYFEPSTFCILLRWALLGFATVYGT 498
Db 615 vlyrdtpvkvssrelcyiilagiflygvcpftliaqpttscylqrllvlgssamcysa 674
QY 499 VTLKLRVLVKV-----FLSRTAQRIPYTGGRVMRMLAVILLVWFVFLIGW 544
Db 675 lvtktnrtarilagskklctkrprfmsawaqvilasillsvqltlvvtlii----- 726

```

CC intracellular (ICD) domains, each chosen independently from a Car
CC (gamma-aminobutyric acid receptor); (2) an optional linker attached to
CC the C-terminus of ICD; and (3) a G-protein (GP) linked to ICD or the
CC linker. (1), and recombinant chimeric receptors (CR) without the GP
CC component, are used to assess function of the various domains and to
CC identify compounds (e.g. allosteric modulators or antagonists) that act
CC on these domains. The modulators are potentially useful for treating or
CC preventing diseases associated with the receptors, e.g. stroke, head or
CC spinal cord injury, epilepsy, ischemia, hypoglycemia, anoxia,
CC Alzheimer's disease, hyperparathyroidism, osteoporosis, cognitive
CC disorders and depression. Nucleic acid (II) that encodes (I) is used:
CC (1) for recombinant production of corresponding proteins; and (2) to
CC produce cells used in screening for modulators. Use of Car and mGluR
CC domains allows presentation of GABAR domains, to a binding agent, in a
CC form more like the natural domain structure compared with use of
CC incomplete receptors, lacking one or more domains. By shuffling different
CC domains, agents can be identified that affect particular domains of a
CC receptor.

XX Sequence 1276 AA;

Query Match 2.9%; Score 183.5; DB 20; Length 1276;

Best Local Similarity 21.8%; Pred. No. 0.00016;

Matches 111; Conservative 74; Mismatches 203; Indels 121; Gaps 24;

QY 311 QKVDIDQCSDDGWFSGTHKC---HLNNSC-CMPKIGLGVFLGAYECICRAG---FYHGGV 363

DB 475 eqvtfdcc---gdvgnyslinwhlspedgsvfkevgv---ynvyakkgerlfincek 527

QY 364 LPVNNFRRRGPDQH-----ISGSTKDVSEE-----AYVCLPCRREG-----CPFCAD 404

DB 528 ilwsgfsrevpfncsrdclagrkilgeptccfecvcpdgeysdetdasackncpd 587

QY 405 D-----SPCFVQEDKYLR-----LAISFGQLCMLLDFFVSMVYHYHFRKAKSIR 448

DB 588 dfwsnennhtscfelpqeyirgdawavgvptiaclgalatl---fvlgvfrh-natpvvk 644

QY 449 ASGLILETILFGLSLYPPVVTLYPEPSTFCILLRWALLGFAFVGVTVTLKLRVLK 508

DB 645 asgrelyllggvfyicymctffiaakpstavctlrllglgtafsvcsalltktnr 704

QY 509 VF--LSRTAQIPYMT--GGVMMMLAVI---LLVVFWEFI---GWTSSVQNLEKQISL 558

DB 705 ifggarregagrprfispasqvaiclaaligsgqllivvawlvveapgtgketapevrv 764

QY 559 ICGGKTSD---HLIFNMCLIDRWYMTVAEFLFLMGVYLCYAVTVPSAFHEPRYMAV 615

DB 765 rchrdasmglslaynvlil-----aictly-afntrkcpenfneakfigf 809

QY 616 AVHNELIISAIFHTIRFVLAS--RLQSDWMLMLYFAHTHTLTVTITGLLLIPKF----- 667

DB 810 tmyttciilwllalpifvrtssyrvqvtmcv-----svslsgsvlglcfapklhllfq 865

QY 668 -----SHSNPNRDIATAYEDELDMRGSGSY-----LNSSIN 701

DB 866 pqknvshraptrfsgsaarssllgqsgsfvptvcngrevvdstsslmtleslma 925

QY 702 SAWSEHSLOPEDIRDELKLYAQLEIYKR 730

DB 926 ccalseakearindeleir---qlrrdrk 951

RESULT 14

AAR25081

ID AAR25081 standard; Protein; 906 AA.

XX

AC AAR25081;

XX

DT 04-JAN-1993 (first entry)

XX

DE GLU-G-R subtype 1b.

XX GLU-G-R; Xenopus oocyte; cerebellum sublibrary; voltage clamp assay;
KW neurotransmitter; glutamate; neuronal excitation.

XX Synthetic.

XX WO9210583-A.

XX 25-JUN-1992.

XX 12-DEC-1991; 91WO-US09422.

XX 12-DEC-1990; 90US-0628806.

XX 30-JAN-1991; 91US-0648481.

XX 18-MAR-1991; 91US-0672007.

XX (UNIW) UNIV WASHINGTON.

XX (ZYMO) ZYMOGENETICS INC.

XX

XX Almers W, Hagen FS, Houamed KM, Mulvihill ER;

XX WPI: 1992-234638/28.

XX N-PSDB; AAQ25815.

XX Mammalian G protein-coupled glutamate receptors - activate

XX phospholipase C or stimulate inositol phospholipid metabolism,

XX for use in diagnosis and identification of receptor agonists and

XX antagonists

XX Disclosure; Fig 7; 144pp; English.

XX

CC The sequence given is mammalian G protein-coupled glutamate

CC receptor (GLU-G-R) subtype 1b. The clone encoding this protein was

CC isolated from Xenopus oocyte cultures which had been injected with RNA

CC extracted from a cerebellum sublibrary. Expression of GLU-G-R

CC activity was assessed by voltage clamp assay. Plasmid DNA encoding

CC GLU-G-R activity was isolated by replica-plating experiments. The

CC protein encoded by clone 45-A was found to act as a receptor for the

CC neurotransmitter glutamate, and may play a critical role in the control

CC of neurons. GLU-G-R and antibodies may be used to identify agonists

CC and antagonists of GLU-G-R-mediated neuronal excitation and in

CC diagnosis.

XX

XX Sequence 906 AA;

Query Match 2.8%; Score 180; DB 13; Length 906;

Best Local Similarity 20.6%; Pred. No. 0.00017;

Matches 93; Conservative 66; Mismatches 170; Indels 122; Gaps 15;

QY 357 GFYHFGVLPVNNFRRRGPDQHSSTKDVSEE-----AYVCLPCRREG- 398

DB 498 glwhegvlnidvdkq---mnksgmrvscvpcclkgqkvirkgevcswcictackene 554

QY 399 -----CPFC-----ADDSPCFVQEDKYLR-----LAISFGQLCMLLDFFVSMV 438

DB 555 fvqdeftcracdlgwwphaeltgcepiprvylewsdiesilaiafscglvltvltlf 614

QY 439 YHFRKAKSTRASGLILLETILFGLSLLYFPVVLVYFESTFRCILLRWALLGFAFVTVGT 498

DB 615 vlyrdtpvkvssrelcyllagiflgyvcpftliakpttscylqrlivgissamcysa 674

QY 499 VTLKLRVLKLV-----FLSRTAQIRIPVMTGGVRVRLAVTLVLVVFVFLIGW 544

DB 675 lvtktnrliarilagakkictrkprfmsawaqviiasllisvqltlvltlii----- 726

QY 545 TSSVCONLEKQISLIGOGKTSIHLIFNMCLIDRWYMTVAE-EFLFLMGVYLCYAVTV 603

DB 727 -----mepmpilyspsike--vyllicntslngvavpvgngllmsectyafakttrv 777

QY 604 PSARFHEPRYMAVAVHNELIISAIFHTIRFVLASRLQSDWMLMLYFAHTHTLTVTITGL 663

DB 778 panfneakyiaftmyttctciwlafvlyf-----gsnykiiitcfavslsvlalgcmf 831

QY 664 IPK-----FSHSSNPRDDIATEAYEDELDMG-----RSGSYLNSINSINSAWSEHSLDPED 713
Db 832 tpkmyiiaakpernvrsafitsdvv-rmdvgdglpcrsntfln----- 874
QY 714 IRDELKLYAQLIYKRRKMTNNPHLQKKR 744
Db 875 -----ifrrkkpgagn---akkr 889

RESULT 15
AAB42491
ID AAB42491 standard; Protein; 3266 AA.
XX AAB42491;
XX DT
XX DE
XX DE

08-FEB-2001 (first entry)
Human ORFX ORF2255 polypeptide sequence SEQ ID NO:4510.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
hypotensive; dermatological; immunosuppressive; antiinflammatory;
antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
neurodegenerative disorder; osteoarthritis; graft vs host disease;
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
cholesterol ester storage; systemic lupus erythematosus; infection;
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
bone damage; cartilage damage; antiinflammatory disease; coagulation;
thrombosis; contraceptive.

Homo sapiens.
XX OS
XX PN W0200058473-A2.
XX PD
XX PF 05-OCT-2000.
XX PF 31-MAR-2000; 2000WO-US08621.
XX PR 31-MAR-1999; 99US-0127607.
XX PR 02-APR-1999; 99US-0127636.
XX PR 05-APR-1999; 99US-0127728.
XX PR 30-MAR-2000; 2000US-0540763.
XX PR (CURA-) CURAGEN CORP.
XX PI Shimkets RA, Leach M;
XX DR WPI; 2000-602362/57.
XX DR N-PSDB; AAC76700.
XX XX
XX PT Novel nucleic acids and peptides derived from open reading frame X,
XX PT useful for treating e.g. cancers, proliferative disorders,
XX PT neurodegenerative disorders and cardiovascular disease -
XX PS Claim 11; Page 3700-3708; 5507pp; English.
XX XX
XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
XX CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX CC antidiabetic; hypotensive; dermatological; immunosuppressive;
XX CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX CC antithyroid; and antianaemic. The sequences can be used for determining
XX CC the presence of or predisposition to, or preventing or treating
XX CC pathological conditions associated with an ORFX-associated disorder. The
XX CC nucleic acids can be used to express ORFX proteins in gene therapy

CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX SQ Sequence 3266 AA;

Query Match 2.8%; Score 179.5; DB 21; Length 3266;
Best Local Similarity 20.8%; Pred. No. 0.0013;
Matches 130; Conservative 78; Mismatches 250; Indels 167; Gaps 22;

QY 659 HSSNNPRDDIATEAYEDELDMGRSGSYLNSINSINSAWSEHSLDPED-IRDELKLYAQL- 726
Db 985 hkspr---asalye-----ssrlsflldredkrlrderlssler 1025
QY 727 ----IYKRKKMTNNPHLQKKRCKGKGLGRSIMRRIPEITVSRQCKEDKEGADHGTA 782
Db 1026 nkfysfaldtkitpdtkalleraaksissreenwsfldwsrfanfnnkdkvdsapr 1085
QY 783 KGTALIRKNPPESGNTGK-----SKEETLKNRVFSLKKSHT-----YDHVRD 826
Db 1086 pipswymkkkirtsdsegkmdkdkhkeeqerqelfasrfihssifedskrlqlher 1145
QY 827 QTESSSLPTFSEEEETE--NSTLES-----SGKKLTQKLEDS 865
Db 1146 keedsdflsgrlygkqtseganstdsiqepvlfhsrfmeltmqdkkkekdkpkve 1205
QY 866 EAESTESVPLVCKSASAHNLSSEKKT----GHPRTSMLOKSLSVIASAKETLC----- 915
Db 1206 kqedtenhpktpesa-penkdselktpsvgppsvtvv--tleaspsalektigtktvea 1262
QY 916 -LAKTOT-----AGVEERTKS 931
Db 1263 plvteektvpatvseakpasepapapvegleqvdipgagdpkeaaampagveegssg 1322
QY 932 QKPLPKD-KETNRNHSNDNTETKDPAQNSNPAEPRKPKQSGIMQQRVNPTTANSDL 990
Db 1323 dpppyldakptpgassfsgaesnvdpdpdatqplskp--aqkseeaneppkaekpdada 1380
QY 991 NPGTTQMKDNFDIGEVCPWEVYDLTPGPVPSEKQKHVSIVASEMEKNPTFSLKEKSHH 1050
Db 1381 epdanqka-----eaapesqpase-----dlevdpvvaakdkkpn 1416
QY 1051 KPAAEVCQOSNOKRIDKAEVCLWESQGSILEDEKLLISKTP-----VLPERAKEENG 1105
Db 1417 kskrsktpvqaaavsiavekpv---trkseridrekllrnsprgeaqkillekmeaek1 1472
QY 1106 QPRAANVCAGQSEELPPKAVASKTENENLQI----GHQEKKTSSSEENVGVSNSNNF 1161
Db 1473 trtasksaadlnehpepslplstrirrvsvyatmgdhenr-spvkepveqprvtrkl 1531
QY 1162 QOPLTSRAEVCPEFETPPAOPNAGR 1186
Db 1532 erelqeaav-----pttprrgr 1549

Search completed: November 1, 2001, 10:14:19
Job time: 3712 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 08:59:27 ; Search time 26.44 Seconds
(without alignments)
946.190 Million cell updates/sec

Title: US-09-775-181-2
Perfect score: 6382
Sequence: 1 MGAMAYPLLLLLLAQLGLG.....LSANKIAGPRKEIWDSPKV 1215

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	220.5	3.5	1199	1	US-08-041-538-2
2	220.5	3.5	1199	1	US-08-463-642-2
3	220.5	3.5	1199	1	US-08-455-602-2
4	220.5	3.5	1199	2	US-08-465-157-2
5	220.5	3.5	1199	5	PCT-US91-09422-2
6	220	3.4	1219	2	US-08-687-289A-6
7	204	3.2	1180	4	US-08-660-148-2
8	200	3.1	1180	3	US-08-486-270-8
9	200	3.1	1180	3	US-08-367-264-8
10	197.5	3.1	1212	4	US-08-660-148-5
11	193.5	3.0	1212	3	US-08-486-270-10
12	193.5	3.0	1212	3	US-08-367-264-10
13	193	3.0	1180	3	US-08-072-574-8
14	190	3.0	1056	2	US-08-687-289A-8
15	186.5	2.9	1212	1	US-08-072-574-10
16	186	2.9	1056	2	US-08-687-289A-7
17	178	2.8	571	4	US-08-961-083-4
18	177	2.8	906	1	US-08-486-270-2
19	177	2.8	906	3	US-08-367-264-2
20	174	2.7	906	5	PCT-US91-09422-17
21	173.5	2.7	1964	2	US-08-790-912-3
22	173.5	2.7	2052	2	US-08-790-912-2
23	172.5	2.7	1079	1	US-08-485-588-8
24	172.5	2.7	1079	1	US-08-484-565-8
25	172.5	2.7	1079	2	US-08-480-751-8
26	172.5	2.7	1079	2	US-08-943-986-8
27	172.5	2.7	1079	3	US-08-353-784-8

Query Match 3.5% Score 220.5; DB 1; Length 1199;

ALIGNMENTS

RESULT 1
US-08-041-538-2
; Sequence 2, Application US/08041538
; Patent No. 5385831
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R
; APPLICANT: Hagen, Frederick S
; APPLICANT: Houamed, Khaled M
; APPLICANT: Almers, Wolfhard
; TITLE OF INVENTION: G PROTEIN COUPLED GLUTAMATE
; TITLE OF INVENTION: RECEPTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: Steuart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA: US/08/041,538
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/648,481
; FILING DATE:
; APPLICATION NUMBER: US 07/626,806
; FILING DATE: 12-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 13952-6-1
; TELEPHONE: 206-467-9600
; TELEFAX: 206-623-6793
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1199 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-041-538-2

Sequence 8, Appli
Sequence 12, Appl
Sequence 12, Appl
Sequence 6, Appli
Sequence 6, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 12, Appl
Sequence 12, Appl
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 26, Appli

28 172.5 2.7 1079 3 US-08-484-719B-8
29 170 2.7 877 1 US-08-486-270-12
30 170 2.7 877 3 US-08-367-264-12
31 168.5 2.6 1848 4 US-08-296-791-6
32 168.5 2.6 1848 5 PCT-US95-10661A-6
33 168 2.6 1863 2 US-08-603-753D-2
34 168 2.6 1863 4 US-09-099-753-2
35 168 2.6 1863 4 US-08-986-106-2
36 165 2.6 877 1 US-08-072-574-12
37 163.5 2.6 905 1 US-08-072-574-2
38 162.5 2.5 872 3 US-08-337-797A-2
39 162.5 2.5 872 3 US-09-258-523-2
40 162.5 2.5 879 1 US-08-072-574-6
41 162.5 2.5 879 1 US-08-486-270-6
42 162.5 2.5 879 3 US-08-367-264-6
43 162 2.5 1780 1 US-08-769-309A-5
44 162 2.5 1780 3 US-08-994-570-5
45 160.5 2.5 1210 3 US-08-545-860D-26

```
Best Local Similarity 18.6%; Pred. No. 2.1e-09;
Matches 163; Conservative 120; Mismatches 330; Indels 265; Gaps 29;

QY 357 GFYHGVLPVNNRRRGGPDQHTSGSKDYSEE-----AYVCLPCREG- 398
Db 498 GTWHEGVNLIDDKYIQ---MNSGMVRSVCSEPCCLKGQIKVIRKGEVSCCWTACKENE 554
QY 399 -----CPFC-----ADDSPCFVQEDKYLK-----LAISFQGLCMLLDFVSMVLV 438
Db 555 FVODEFTCRACDLGWWPNAELTGCCEPIPVRYLEWSDIESIIAIAFSCGLILVTLFVTLIF 614
QY 439 YHFRKAKSTRASGLILLETILFGLSLLFPVVLVYFEPSTFCILLRWALLGFATVYGT 498
Db 615 VLYRDPVVKSSRELXYILLAGIFLGYVCPPTLAKPTTSCYLQRLVLGSSAMCYS 674
QY 499 VTLKLRVLKV-----FLSRTAQRIPYMTGGRVWRMLAVILLVVFVFLIGW 544
Db 675 LVTKTNRIARILAGSKKICTRKRFRMSAWAQVIAIISVQLTLVVTLLI----- 726
QY 545 TSSVCONLEKQISLIGOGKTSDHILFNCLIDRWDMYMTAVA-EFLFLMGVYLCYAVRTV 603
Db 727 -----MEPPMILSYPSIKE--VILICNTSNLGVAPVGYNGLLIMSCITYAFKTRNV 777
QY 604 PSAPHEPRYMAVAVHNELIISAIFHTIRFVLASRLQSDWMLMLYFAHTHLTVTVTIGLL 663
Db 778 PANFNEAKYIAFTMTTCTIWLAFVPIVF-----GSNYKIIITTCFAVSLSVTVLGCMF 831
QY 664 IPK-----FSHSSNPRDDIATAYEDELDMG-----RSGSYLNSSINSANSSEHSLDP 711
Db 832 TPKMYIIIAKPERNVRSAPTT---SDVVRMHVGDGKLPCRSTNFL----- 874

QY 357 GFYHGVLPVNNRRRGGPDQHTSGSKDYSEE-----AYVCLPCREG- 398
Db 498 GTWHEGVNLIDDKYIQ---MNSGMVRSVCSEPCCLKGQIKVIRKGEVSCCWTACKENE 554
QY 399 -----ADDSPCFVQEDKYLK-----LAISFQGLCMLLDFVSMVLV 438
Db 555 FVODEFTCRACDLGWWPNAELTGCCEPIPVRYLEWSDIESIIAIAFSCGLILVTLFVTLIF 614
QY 439 YHFRKAKSTRASGLILLETILFGLSLLFPVVLVYFEPSTFCILLRWALLGFATVYGT 498
Db 615 VLYRDPVVKSSRELXYILLAGIFLGYVCPPTLAKPTTSCYLQRLVLGSSAMCYS 674
QY 499 VTLKLRVLKV-----FLSRTAQRIPYMTGGRVWRMLAVILLVVFVFLIGW 544
Db 675 LVTKTNRIARILAGSKKICTRKRFRMSAWAQVIAIISVQLTLVVTLLI----- 726
QY 545 TSSVCONLEKQISLIGOGKTSDHILFNCLIDRWDMYMTAVA-EFLFLMGVYLCYAVRTV 603
Db 727 -----MEPPMILSYPSIKE--VILICNTSNLGVAPVGYNGLLIMSCITYAFKTRNV 777
QY 604 PSAPHEPRYMAVAVHNELIISAIFHTIRFVLASRLQSDWMLMLYFAHTHLTVTVTIGLL 663
Db 778 PANFNEAKYIAFTMTTCTIWLAFVPIVF-----GSNYKIIITTCFAVSLSVTVLGCMF 831
QY 664 IPK-----FSHSSNPRDDIATAYEDELDMG-----RSGSYLNSSINSANSSEHSLDP 711
Db 832 TPKMYIIIAKPERNVRSAPTT---SDVVRMHVGDGKLPCRSTNFL----- 874

QY 712 EDIRDELKLYAQLETKYKRRKMTNPNHLOKRC-----KGLGRSIRMRITRIPET 764
Db 875 -----IFRKKKAGNANSKSVSWSEPGGROAPKQGHVQRLSVHVKT 919
QY 765 VSRQSKDEKAGDHCTAGTALIRKNPSSGNTGKSEETLKNRVFSKKSHSTYDHY 824
Db 920 NETACNQ-----TAVIK---PLTKSYQSGSKSLTFSD----- 948
QY 825 RQTEBSSSLPTESQEBETSTLESKGLTKQKLKEDSEASTESVPLVCKSASAHN 884
Db 949 -----ASTKLYNVEEDNTPSAHFSPSPSPSMVVR 980
QY 885 LSSEKKTGHPTSMLOKLSVSIASAKETLGLAGKTQTAGVEERTKSKQPLPKDRETNRN 944
Db 981 RGPVATTPPLPHL-----TAEETPLFADSVIPKGL-----PPLPQQO----- 1021
QY 945 HNSDNTETKDPAPQNSPAEPRKPKQSGIMKQQRVNPTTANSDL-----NPGTQ 996
Db 1022 -----POQPPQPPQPPQPK--SLMDQLQGVVTFNFGSGIPDFHVLAGPGTP- 1065
QY 997 MKDNFDIGEVCPWEVYDLTPGPVPSE--SKVOKHVSIVASEMEKNPTFSLKEKSHHKPKA 1054
Db 1066 -----GN-----SURLYPPPPPHLQMLPLHUSTFOEESISPPGEDIDDD 1108
QY 1055 AEVCOOSNOKRIDKAEVCLWESQG---QSILEDKILLISKTPVLPRAKEENGOPRAAN 1111
Db 1109 -----ERKLLQEFVYEREGTEDEDEEEDLPTASKLTPEDSPALTPPSPFRDS 1159
QY 1112 VCAGOSEELPP--KAVASKTENENLNQHQHEKQYSS 1147
Db 1160 VASGSSVSPSPSVESVCLTPPNVTYASVILRDYKQSSS 1197

RESULT 2
US-08-463-642-2
; Sequence 2, Application US/08463642
; Patent No. 5721107
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R
; APPLICANT: Hagen, Frederick S
; APPLICANT: Houamed, Khaled M
; APPLICANT: Alimers, Wolfhard
```


Db 1160 VASGSSVPSPVSESVLCTPPNVTYASVILRDYKQSSS 1197

RESULT 4

US-08-465-157-2
; Sequence 2, Application US/08465157
; Patent No. 5869609
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R.
; APPLICANT: Hagen, Frederick S.
; APPLICANT: Houamed, Khaled M.
; APPLICANT: Almers, Wolfhard
; TITLE OF INVENTION: G PROTEIN COUPLED GLUTAMATE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; CITY: Steuart Street Tower, One Market Plaza
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,157
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/041,538
; FILING DATE:
; APPLICATION NUMBER: US/07/648,481
; FILING DATE:
; APPLICATION NUMBER: US 07/626,806
; FILING DATE: 12-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 13952-6-1
; TELEPHONE: 206-467-9600
; TELEFAX: 206-623-6793
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1199 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-465-157-2

Query Match 3.5%; Score 220.5; DB 2; Length 1199;
Best Local Similarity 18.6%; Pred. No. 2.1e-09;
Matches 163; Conservative 120; Mismatches 330; Indels 265; Gaps 29;
Qy 357 GFYHPCVLVNNFRFRGPDQHSIGSTKDVSEE-----AYVCLPCREG- 398
Db 498 GTWHGVLNIDDKYIQ-----MNKSGMVRVSCPELKGQIKVIRKGEVSCWICTACKENE 554
Qy 399 -----CPFC-----ADSPCFQVEDKYL-----LAISFQGLCMLLDVFSMLV 438
Db 555 FVQDEFTCRACDLGWNPAELTGCEPIPVRYLEWSDISIAIAFSCGLVLTFLVTLIF 614
Qy 439 YHFRKAKSIRASGLLITLFGSLLLPVYVFPSTFRCILLRWARILGFATVYGT 498
Db 615 VLRYDTPVVKSSRELGYILLAGIFGVCPFTLIAKPTTTSCYLQRLVLGLSSAMCYSA 674
Qy 499 VTLKLRVLKV-----FLSRTAQRIPYMTGGRVMRLAVILLVWFELIGW 544

Db 675 LVTKNRIARIAGSKKKICTRKRPRFMSAWAQVILIASILSVOLTLVVTLLI----- 726
Qy 545 TSSVCONLEKQISLQOGKTSDLHIFNMCLIDRWYMTAVA-EFLFLWGVYLCVAVRTV 603
Db 727 -----MEPPMPILSYSIKE--VYLICNTSNLGVAPVGVNGLLIMCTYYAKTRNV 777
Qy 604 PSAPHEPRYMAVAVHNELIISAIFHTIRFVLASRIQSDMMLMLYFAHTHLTVTVTIGLL 663
Db 778 PANENEAKYIAFTMTTCIWLAFVPIV-----GSNYKIITTCFAVSLSVTVALGCMF 831
Qy 664 IPK-----FSHSSNPRDDIATEAYEDELDMG-----RSGSYLNSNSINSAWSEHSLDP 711
Db 832 TPKMYIIIAKPERNVSFTT---SDVVRMHVGDGKLPCRSNTFLN----- 874
Qy 712 EDIRDELKLYAQLIYKRRKMITNNPHLQKKRCS-----KGLGRSIRMRIITEIPET 764
Db 875 -----IFRRKKPGAGNANSKSVSWSEPGGQAPKQGHVWQRLSVHVKT 919
Qy 765 VSRQCKSKEDKAGDHGTAKGTALIRKNPPSSGNTCKSKEETLKNRVFSLKKSHTYDHW 824
Db 920 NETACNQ-----TAVIK---PLTKSYOGSGKSLTFSD----- 948
Qy 825 RDQTESSSLPTESQEBEETTENSTLESLSGKLTOKLKEDSEASTESVPLVCKKSASAHN 884
Db 949 -----ASTKTLYNVEEDNTPSAHFSPSPSPSPSPSPSPSPSPSPSPSPSPSP 980
Qy 885 LSSEKKTGHPRTSMLOKSLSVIASAKETLGLAGKTQTAGVEERTKSKOKPLPKDKETNRN 944
Db 981 RGPVATTPPLPHL-----TAEETPLFADSVIPKGL-----PPPLPQQO----- 1021
Qy 945 HNSDNTETKDPAPQNSNPAPBPRKQKSGIMKQKRVNPTTANSDL-----NPGTTQ 996
Db 1022 -----PQPPPPQPPQPPK--SLMDQLQGVVTFNGSGIPDFHVLAVLAGPTP- 1065
Qy 997 MKDNFDIGECVPEVYDLTPGPVPS--SKVQKHVSIIVASEMEKNTPTSLKESKSHKPKA 1054
Db 1066 -----GN-----SLRSLYPPPPPPQHLMLPLHLSTFQEEISIPPGEDIDDD 1108
Qy 1055 AEVCOQSNQKRIDKAEVCLWESOG---OSILEDEKLILSKTPTLPERAKEENGOPRAAN 1111
Db 1109 -----ERFKLQEFYEREGNTEDELEEDLPTASKLTEDSPALTPPSPFRDS 1159
Qy 1112 VQAGOSELPP--KAVASKTENENLNQIGHOEKKTSS 1147
Db 1160 VASGSSVPSPVSESVLCTPPNVTYASVILRDYKQSSS 1197

RESULT . 5
PCT-US91-09422-2
; Sequence 2, Application PC/TUS9109422
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R.
; APPLICANT: Hagen, Frederick S.
; APPLICANT: Houamed, Khaled M.
; APPLICANT: Almers, Wolfhard
; TITLE OF INVENTION: G PROTEIN-COUPLED GLUTAMATE RECEPTORS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: One Market Plaza, Steuart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/09422
; FILING DATE: 19911212
; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA: US 07/672,007
;; FILING DATE: 18-MAR-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/648,481
;; FILING DATE: 30-JAN-1991
;; PRIOR APPLICATION DATA: US 07/626,806
;; FILING DATE: 12-DEC-1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Parmelee, Steven W.
;; REGISTRATION NUMBER: 31,990
;; REFERENCE/DOCKET NUMBER: 13952-6PC
;; TELEPHONE: (206) 467-9600
;; TELEFAX: (415) 543-5043
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1199 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
PCT-US91-09422-2

Query Match 3.5%; Score 220.5; DB 5; Length 1199;
Best Local Similarity 18.6%; Pred. No. 2.1e-09;
Matches 163; Conservative 120; Mismatches 330; Indels 265; Gaps 29;

QY 357 GFYHPGLVNNFRPPDOHSGSTKDVSEE-----AYVCLPCREG-398
Db 498 GTWHEGVNIDDDYKIQ---MNKSMVRSCVSEPCLKQIKVIRGEVSCCMWITACKENE 554
QY 399 -----CPFC-----ADDSPCFQVEDKYLR-----LAITSFOGLMCLDFVSLVV 438
Db 555 FVQDEFTCRACDLGWNFAELTGEPTPVRYLEWSDIESIIATFSCGLIIVTLVFLIF 614
QY 439 YHPRKAKSIRASGLILETILFGSLLYFPVILYFPSTFRCLLRLKARLLGFATVYGT 498
Db 615 VLYRDTPWKSSRELCCYIILAGIFLGYVCPFTLIARPTTSCYLRLLVGLSSAMCYSA 674
QY 499 VTLLKHLRVLVK-----FLSRTAORIPYMTGGVRMRLAVILLVFWFVLIGW 544
Db 675 LVYKTNRIARILAGSKKIKTRPRFMSAWAQVFIASILISQVTLVTLII----- 726
QY 545 TSSVQCNLEKQISLIGQKTSIDLIFNMCLIDRWYMTAVA-EPLFLWGVYLCYAVRTV 603
727 -----MEPPMPLSYPSIKE--VYLICNTSLNGVVPVVCYNGLLIMSCITYYAFKTRNV 777
QY 604 PSAPHEPRYAVAVHNELIISAIPHTIRFVLASRLQSDWMLMLYFAHTHTVTVTIGILL 663
Db 778 PANFNEAKYIAFTMYTTCIIWLAFPVYF-----GSNYKIITTCFAVSLSVTVALGCMF 831
QY 664 IPK-----FSSHNNPRDDIATEAEVDELDMG-----RSGSVLNSINSANSEHSLDP 711
Db 832 TPKMYIIIAKPERNVSFAFT---SDVVRMHVGDGKLPCRSNTFLN----- 874
QY 712 EDIRDELKLYAOEIKYKRMKMITNPHLOKRCs-----KKGLGRSMRRITETIPET 764
Db 875 -----IFERRKPGAGNANSNCKSVSWSEPGGRQAPKQGHVWVORLSVHVKT 919
QY 765 VSRQCKEDKAGDHGTAKTALIRKKNPPSSNGTGSKEETLKNRVFLSKKSHSYDHY 824
Db 920 NETACNQ-----TAVIK---PLTKSYQSGSKSLTFSD----- 948
QY 825 RDQTESSSLPTESQEBETTENSTLESGLKLTQKLKEDSEABESTESVPLVCKSAHN 884
Db 949 -----ASTKTLYNVEEDNTPSAHFPSPSPSPSVVHR 980
QY 885 LSSEKKTGHPRTSMLOKSLSVIASAKETLGLAGKTQTAGVEERTKSOKPLPKDKETNRN 944
Db 981 RGPVATTPPLPHL-----TAETPLFLADSVIPKGL-----PPPLPQQQ----- 1021

QY 945 HNSNDTETKDPQNSNPABEPRKPKSGIMKQORVNPPTANSDL-----NPGTTQ 996
Db 1022 -----PQPPPPQPPQPPK--SLMDQLQGVVTFNGSGIDFHFHVLAPGPTP- 1065
QY 997 MKDNFDIGEVCPMEVYDLTGPVPSE--SKVOKHVSVIVASEMEKNPTFSLKEKSHHKPKA 1054
Db 1066 -----GN-----SLRSLYPPPPPPQHLQMLPLHLSTFQESISPPGEDIDDS----- 1108
QY 1055 AEVCQSQNKRIDKAEVCLMESQG---OSILEDEKLISKTPVLPRAKEENGOGPRAAN 1111
Db 1109 -----ERFKLLQEFVYVEREGTEDELEEDLPTASKLTPEDSPALTPSPFRDS 1159
QY 1112 VCACQSEELPP--KAVASKTENENLNOIGHQEKTKTSS 1147
Db 1160 VAGSSVPSPPSVSESVLCTPPNVTYASVILRDYKQSS 1197

RESULT 6
US-08-687-289A-6
; Sequence 6, Application US/08687289A
; Patent No. 5981195
; GENERAL INFORMATION:
; APPLICANT: Fuller, Forrest H.
; APPLICANT: Krapcho, Karen J.
; APPLICANT: Hamerland, Lance G.
; TITLE OF INVENTION: CHIMERIC RECEPTORS AND METHODS FOR
; IDENTIFYING COMPOUNDS ACTIVE AT
; TITLE OF INVENTION: METABOTROPIC GLUTAMATE RECEPTORS AND
; TITLE OF INVENTION: THE USE OF SUCH COMPOUNDS IN THE
; TITLE OF INVENTION: TREATMENT OF NEUROLOGICAL DISORDERS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Fastseq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,289A
; FILING DATE: July 25, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,526
; FILING DATE: July 26, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 220/004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1219 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-687-289A-6

Query Match 3.4%; Score 220; DB 2; Length 1219;
Best Local Similarity 18.8%; Pred. No. 2.3e-09;


```
QY 556 ISL-----IGOKTSDHLIFNMCLIDRDWYMTAVA-BEFLFLMGVYLCYAVRTVPSAF 607
Db 709 VALFIMEPPDHMDYPSIREYVILCTNTLGVVTPGLYNGLLILSCTFYAFKTRNVPANF 768
QY 608 HEPYMAVAVHNELIIISAIFHTIRFVLASRLQSDWMLMFAHTHLTVTTIGLLIPK- 666
Db 769 NEAKYIAFTMYTTCIIWLAFVPIYF-----GSNYKIITMCFVSLSATVALGCMFVKV 822
QY 667 ---FSHSSNNPRDDIATE-AYEDELMDGRSGSYL--NSSINSAMSEHSGLDPDIRDELK 720
Db 823 YIILAKPERNVRSAFTTSTVVRMHVGDGKSSAASRSSLVNLWKRRGSSGETLSSNGKS 882
QY 721 -----LYAQLEIYKRRKMTNNPHLOK---KRCSSKGLGRSIRMRIPE 763
Db 883 VTWAQNEKSRGQHLWQRLSIHINKKENPNQTAIVKPFKSTESRGLGAG----- 932
QY 764 TVSRQSKEDKEGADHTAGTALIRKNPPESGNTGKSKEETLKNRVFSLKKSHSTYDH 823
Db 933 -----ACAG-GSAGGVGATGAGCAGAGPGPESPDAGPKALY----- 969
QY 824 VROTESSSLPTESQEEETTENTLESGLKKLTOKLKEDSEAEATESVPLVCKSASAH 883
Db 970 --DVAEAEHFAPARPRSPSTLSHRAG---SASRTDDDDVPSLHSEPVARSSSOG 1023
QY 884 NLSEKKTGHPR-TSMLQKSLSVIASAKEKTLGLAG-----KTQTAGVEE 927
Db 1024 SLMEQISSVVYTRFANISELNSMLSTAAPSPGVGAPLCSSYLIPKEIQLPTMTTFAE- 1082
QY 928 RTRSQKPLPKDKETNRNHSNDTETKDPAPON--SNPAEPRKPKQSGIMKQORVNPTT 985
Db 1083 ----IQPLPAIEVTCGAQPAAGAAQADARESAPGAPEAAAKPDLEIVALTPPSP-- 1136
QY 986 ANSDLNPGTTQMKDNFDIGEVPEWYVYDLTPGPVPSESK 1024
Db 1137 -----FRDSVDGSGTTPNSPVSESALCIPSSPK 1164

RESULT 8
US-08-486-270-8
; Sequence 8, Application us/08486270
; Patent No. 5807689
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; APPLICANT: Johnson, Edwin C.
; APPLICANT: Hess, Stephen D.
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,270
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/072,574
; FILING DATE: 04-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
```

```
; REFERENCE/DOCKET NUMBER: FP41 9772
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1180 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-486-270-8

Query Match 3.1%; Score 200; DB 1; Length 1180;
Best Local Similarity 18.9%; Pred. No. 11e-07;
Matches 178; Conservative 116; Mismatches 401; Indels 248; Gaps 33;

QY 219 ETWFHGLRKRWRPHLRGPNQGRGLGHSWRKDCGLGDKSHFKWSPPLYECENSY- 277
Db 333 DVKWFDDYILKLRPETNHRNP-----WQEFWQHR-----FOCRLEAFPOENSKYN 378
QY 278 -----KPGMLV--TLSSAIYGL---QPNLVPERGV-----MKVD 307
Db 379 KTCNSSLTATHHVDQSKMGFVNAIYSMAYGLHNMQMSLCPGAYAGLCDAMKPIDGRKLL 438
QY 308 INLQKVIDQCSSDGFSGTHKCHLNSECMPKGLGVLGAYECICKAGFYHFGVLPVN 367
Db 439 ESLMKTNFTGVSGDTILFDENGSDPGRYEIMNFKMG-----KDYFDYINVGSWD 488
QY 368 NFRRRGPDQHI-----SGSTKDVSEE-----AYVCLPCREG----- 398
Db 489 NGELKMDDEWVSKSNIIRSVCEPCEKQIKVIRKEVSCWCCTPCKENEYVFDEYT 548
QY 399 CPFC-----ADD-SPCFVQEDKYL-----LAIISFOGLCMLLDFVSMVYHERKAK 445
Db 549 CKACOLGSWFTDDLGTGCDLIPVQVLRWGDPDEPTAAVFAACGLGLIATLFTVTVFIIRDT 608
QY 446 SIRASGLILLETILFGSLLLYFPVVILYFEPSTFRICILLRWARLLGFATVYGTVTKLHR 505
Db 609 VKSSSRRELCYIILAGICLGYLCTFCLIAKPKQIYCYQLRIGIGLSPAMSYSALVTAKTRN 668
QY 506 VLKV-----FLSRTAQRIPYMTGGVRMRLAVILLVWVFWFLIGHTVSSVCQN 551
Db 669 IARILAGSKKKICTPKPRFMSACAQ-----LVIAFILI-----CIQ 704
QY 552 LEKOISL-----IGOKTSDHLIFNMCLIDRDWYMTAVA-BEFLFLMGVYLCYAVRTV 603
Db 705 LGTIIVALTPEPPDHMDYPSIREYVILCTNTLGVVTPGLNGLLILSCTFYAFKTRN 764
QY 604 PSAPHEPRYMAVAVHNELIIISAIFHTIRFVLASRLQSDWMLMFAHTHLTVTTIGLL 663
Db 765 PANFPEAKYIAFTMYTTCIIWLAFVPIYF-----GSNYKIITMCFVSLSATVALGCMF 818
QY 564 IPK-----FSHSSNNPRDDIATE-AYEDELMDGRSGSYL--NSSINSAMSEHSGLDPDIR 716
Db 819 VPKYIILAKPERNVRSAFTTSTVVRMHVGDGKSSAASRSSLVNLWKRRGSSGETLSS 878
QY 717 ELKK-----LYAQLEIYKRRKMTNNPHLOK---KRCSSKGLGRSIRMRI 759
Db 879 NGKSVTWAQNEKSRGQHLWQRLSIHINKKENPNQTAIVKPFKSTESRGLGAG----- 932
QY 760 EIPETVSRQSKEDKEGADHTAGTALIRKNPPESGNTGKSKEETLKNRVFSLKKSHS 819
Db 933 -----ACAG-GSAGGVGATGAGCAGAGPGPESPDAGPKALY----- 969
QY 924 TYDHRDQTESSSLPTESQEEETTENTLESGLKKLTOKLKEDSEAEATESVPLVCKS 879
Db 970 -----DVAEAEHFAPARPRSPSTLSHRAG---SASRTDDDDVPSLHSEPVARS 1019
QY 880 ASAHNLSSEKKTGHPR-TSMLQKSLSVIASAKEKTLGLAG-----KTQTAG 923
Db 1020 SSOGSLMEQISSVVYTRFANISELNSMLSTAAPSPGVGAPLCSSYLIPKEIQLPTMTT 1079
```

```
QY      924   GVEERTKSKPLPKDKETNRNIISNDNTETKDPAQN--SNPAAEPRKPQSGIMKQVR  981
          :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      1080   FAE-----IQPLFAIVTGGAPGAAGAAGDAARSPAAQGEAAAAAKPDLEELVALTPP  1134
          :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      982   NPPTANSDLNPGTTQMKNFIDICEVCPWEVYDITPCPVPSSEK  1024
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      1135   SP-----FRDSVDGSGSTTPNSPVSEALCPISSPK  1164
          :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

RESULTS

US-081-367-264-8
; Sequence 8, Application US/08367264
; Patent No. 6001581
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; APPLICANT: Johnson, Edwin C.
; APPLICANT: Hess, Stephen D.
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
;

Query Match 3.1%; Score 200; DB 3; Length 1180;
Best Local Similarity 18.9%; Pred. No. 1.le-07;
Matches 178; Conservative 116; Mismatches 401; Indels 248; Gaps 33;

Qy	219	ETWFHGLRRKWRPHUHRGPNQGPRGLGHSHWRRKDGLGDGDSHKFKWSPPYLECENGSY-	277
	:	: : :	:
Db	333	DVKWFDYYLKLRPETHRNP-----WFQEFMQHR-----FQCRLEAFPQENSRYN	378
	:	: : :	:
Qy	278	-----KGWLVTLSIAIYL---QPNLVPEFRGV-----MKVD	307
	:	: : :	:
Db	379	KTCNSSLTlKTHHVODSKMGFINAIYSMAVGLHNQMISLCPGYAGLCDAMKPIDGRKLL	438
	:	: : :	:
Qy	308	INLQKVIDIQCSSDGWFSCTHKCHLNNSCEMPTIKGLGFVLGAYEICACAGFVHPGPLVN	367
	:	: ! : ! :	:
Db	439	ESLMKTNFTGVSGDITILFDENGSDSPGYIIMNFKEG-----KDYDFYINVGSMD	488
	:	: : : ! :	:

Qy	368	NFRRGPDQHI-----SGSTKDVSEE-----AYVCLPCREG-----	398
Db	489	NGELKMDDDEWSKKNIIRSVCSPECKGQIKIRKGEVSCWCTCTCKENYVFDEYT	548
Qy	399	CPEC-----ADD-SPCFVQEDKYLR-----LAIISQGLCMLLDFVSMVLVYHFRKAK	445
Db	549	CKACQSGSWPTDGLTCDLIPVOYLRWGDPEPIAAVFAVACGLGLATLFTVTVVFIYRDTP	608
Qy	446	SIRASGLILLETILFGLSLLYFPVITLYPEPSTFCRILLRWARLLGFAVTGVTFLKLR	505
Db	609	VKSSSRELCYIILAGICGLYCTFCLIAKPKQIYCYLQRIQIGLSPAMSVSALVTKNR	668
Qy	506	VLKV-----FLSRTAQRIPYMTGGRVNRMLAVILLVVFVFLIGWTSVVCQN	555
Db	669	IARLAGSKKICKTPKPRMSACQ-----LVIAFIL-----CIQ	704
Qy	552	LEKQISL-----ICQKTSDBLTFNMCLIDRWMTAVAF-EFFLLWGVYLCYAVRTV	603
Db	705	LGIIVALFIMEPPDIMHDYPSIREVLYICNTNLGVVTPGLNGNGLLILSCTFYAFKTRNV	764
Qy	604	PSAFHEPRYMAVAHVNELIISAIFHTIRVLAQSODWMLMLYFAHTHLVATVITGLLL	663
Db	765	PANFPEAKYIAFTMYTCTLIWLAFLPIYF-----GSNKKIITMCFVSLSATVALGCMF	818
Qy	564	IPK-----FHSNNPRDDIATE-AYEDELDMGRSGSYL--NSSINSANSEHSLDPEIRD	716
Db	819	VPKVYIILAKPERNVRSAFTTSTVVRMHVGDCKSSAASRSSSLVNLKRRGSSGETLSS	878
Qy	717	ELKK-----LYAQLEIYKRKKMITNPHLOK---KRCSSKKGLGRSIMRRIT	759
Db	879	NGKSVTWAQNEKSSRGOWLQWRLSHINKENPNQTAIVKPPFKPSTESRGJLAG-----	932
Qy	760	EIPETVSROCKEDKGADHGTAAGTALIRKNPPPESSGNTGKSBETLKNRVFSLKKSHS	819
Db	933	-----AGAG-GSAGGVGATGGACAGAGPGGSPDAGPKALY-----	969
Qy	820	TYDHRVDQTESSSLPESQEBETTENSTLESLSKKLUKQLKEDSEABESTESVPLVCKS	879
Db	970	-----DVAEEHFPAPARPRSPISITLSHRAG-----SASRTDDDDVPSLHSEPVARS	1019
Qy	880	ASAHNLSSEKKTCHPR-TSMLOKLSLVTASAKEKTLGLAG-----KTQTA	923
Db	1020	SSQGLMEQISSVVVTRFTANISELNSMMLSTAAPSPGVGAPLCCSYLIPKEIQLEPTMTT	1079
Qy	924	GVBERTKOKPLPKDKETNRHNSNDNETTKDPAQN--SNPAEPRKPQKSGIMLKQOR	981
Db	1080	FAE-----IQPLPAIEVTGGAQAPAGAQAAGDAARESFAAGPAAAKAPDLEELVALTPP	1134
Qy	982	NPTTANSDLNPGTQMKDNFDIGEYCPWEVVDLTGPGVPVSESK	1024
Db	1135	SP-----FRDSVDSGSGTTPNSPVSESLCALCIPSSPK	1164

```

RESULT 10
US-08-660-148-5
; Sequence 5, Application US/08660148
; Patent No. 6211353
; GENERAL INFORMATION:
; APPLICANT: Burnett, J. P.
; APPLICANT: Mayne, Nancy G.
; APPLICANT: Sharp, Robert L.
; APPLICANT: Snyder, Yvonne M.
; TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
; RELATED NUCLEIC ACID COMPOUNDS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285

```

20


```
Db 609 VYKSSRELCLYIAGICIGYCTECLIAKPKQIYCIORIGIGSPAMSYSALVTKNR 668
Qy 506 VLKQ-----FLSRTAQRIPTYMTGGRVMRLAVILLVVFVFLIGWTSSVCQN 551
Db 669 IARILAGSKKICTPKPREMSACQ-----LVIAFILI-----CQ 704
Qy 552 LKQISL-----TQGGTSOHLIFNMCLIDRWDMYATA-EFIFLLMGVYLCYAVRTV 603
Db 705 LGIIVAFIMEPPDIMHDYPSIREVYLICNTNLGVVTPPLGNGLLILSCTEYAFKTRNV 764
Qy 604 PSAEFPRYMVAVHNELIISAFHTIRVFLASRLQSDWMLLYFAHTHLTVTVTIGLLL 663
Db 765 PANFPEAKYIATMTTCKIWLAFVIYF-----GSNYKIITMCFVSLSATVALGCMF 818
Qy 664 IPK-----FSHNNPRDDIATE-AYEDELDMGRSGSYL--NSSINSAMSEHSLDPEDIRD 716
Db 819 VPKVYIILAKPERNVRSATTTSTVVRMHVGDCKSSAARSLLVNLKRRGSSGETLSS 878
Qy 717 ELKK-----LYAOLEYKRRKKMITNNPHLOK-----KRCSSKGLGRSIMRRIT 759
Db 879 NGKSVTWAQNKSSRGQHLWQLSHINKENPNQTAVIKPPKSTESRGLGAG-----932
Qy 760 EIPFVSROCSEKEDGADHGTAKGTALIRKNPPSSGNTGKSKEETLKNRVFSLKKSIS 819
Db 933 -----AGAG-GSAGGVGATGGACACGPGPESPDAGPKALF-----969
Qy 820 TYDHRDOTESSSLPTESQEBETTENSTLSGKLTOKLKEDESEABESTSVPLVCKS 879
Db 970 -----DYAEAEHPAPARPPSPPIRTLSHRAG-----SASRTDDVPSLHSEPVARS 1019
Qy 880 ASAHNLSSKKTKGHRP-TSMLOKSLSVIASAKEKTLGLAG-----KTQTA 923
Db 1020 SSQGSIMEQISSWTRFTANISELSMMLSTAAPSGVGAPLCCSYLPKEIQLPPTMTT 1079
Qy 924 GVEERTKSKPLPKPKETNRNHSNDTETKDPAPON--SNPAEPRKPKOKSGIMKQORV 981
Db 1080 FAE-----IQPLPALEVGTGAQAAGAAAGAAESPAAGPEAAAKPDLEELVALTTP 1134
Qy 982 NPTTANSLNPGTQMKONDFIGVCPMEVYDLTPGVPVSESK 1024
Db 1135 SP-----FRSDVDSGSTTPNSPVSEALCIPSSPK 1164
```

RESULT 14

US-08-687-289A-8
; Sequence 8, Application US/08687289A
; Patent No. 5981195

GENERAL INFORMATION:

APPLICANT: Fuller, Forrest H.
APPLICANT: Krapcho, Karen J.
APPLICANT: Hammerland, Lance G.
TITLE OF INVENTION: CHIMERIC RECEPTORS AND METHODS FOR
IDENTIFYING COMPOUNDS ACTIVE AT
TITLE OF INVENTION: METABOTROPIC GLUTAMATE RECEPTORS AND
IDENTIFYING COMPOUNDS ACTIVE AT
TITLE OF INVENTION: THE USE OF SUCH COMPOUNDS IN THE
TREATMENT OF NEUROLOGICAL DISORDERS
TITLE OF INVENTION: AND DISEASES
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSEQ for Windows 2.0

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,289A
; FILING DATE: July 25, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,526
; FILING DATE: July 26, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 220/004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1056 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-687-289A-8

Query Match 3.0%; Score 190; DB 2; Length 1056;

Best Local Similarity 19.9%; Pred. No. 6.3e-07;

Matches 124; Conservative 97; Mismatches 241; Indels 162; Gaps 20;

```
Qy 357 GFYHPCVLPVNNFRRGDPDQHSIGSTKDYSE-----AYVCLPCREG- 398
Db 498 GTWHEGVINIDYKIQ---MKNKSGVVRVSCSEPCCLKQIKVIRKGEVSCCMCTACKENE 554
Qy 399 -----CPFC-----ADSPCFQVEDKYL-----LAIISFOGCMLLDFVSMLV 438
Db 555 YVQDEFTKACDLGWWPNADLFGCEPIPVRYLEWNIPIIAFASCGIILVTLFVTLIF 614
Qy 439 YHFRKAKSIRASGLILLETILFGSLLLYFPVVILYFEPESTFCILLRWALRGFATVYGT 498
Db 615 VLYRDTPVVKSSRELCTIILAGIFLGVVCPPTLIAPKPTTSCYLRLLVGLSSAMCYSA 674
Qy 499 VTLKLRHLVKV-----FLSRTAQRIPTYMTGGRVMRLAVILLVVFVFLGW 544
Db 675 LVTKTNRARIILAGSKKICTRKPRFMSAWAQVILIASILISVQLTLVTVTLIT----- 726
Qy 545 TSSVCONLEKQISLIGOGKTSDDLHLENMCLIDRWDMYATAVEF---LFLMGVYLCYAV 600
Db 727 -----MEPPMPIL-----SYPSIKEVILCNTNLGVVAPLGVNGLLIMCTTYAFKT 774
Qy 601 RTVPSAFHEPRYMVAVHNELIISAFHTIRFVLASRLQSDWMLLYFAHTHLTVTVTIG 660
Db 775 RNVPAFNEAKYIAFTMTTCLWLAFVIYF-----GSNYKIITTCFAVSLSVTVVAG 828
Qy 661 LLLIPK-----FSHNNPRDDI--ATEAYEDELDM-----GRSGSYL 696
Db 829 CMFTPKMYIIITAKPERNTEIEVRCSTAAAFKVAARATLRRSNVSRKRSSSLGGSTGTP 888
Qy 697 NSSINSAMSEHSLDPEDIRDELKLYLAQLEYKRRKKMIT---NNPHLOKRCSSK---GL 750
Db 889 SSISSSKSNSEDPFPQPERKQOQPLAULTQEQOQOPLTLPQOQSQOQPKQKQVFGS 948
Qy 751 GRSIMKRITETPVSROCSEKEDGADHGTAKGTALIRKNPPSSGNTGKSKEETLKNR 810
Db 949 G-----TVTFLSFLDFEPQKNMAHGNS-----THQNS 975
Qy 811 VFSLKKSHTYDHRDQTESSSLPTESQEBE---TTNNTSLSGSKKLQKLEDESEA 867
Db 976 LEAKSSDTL-----TRHQLPLPQCGETDLDTVQETGLQGPVGGDQRPVEDPEEL 1028
Qy 868 ESTESVPLVCKKSASAHNLSSEKKT 891
Db 1029 ---SPALVVSQSQSFVISGGGST 1048
```


Align seg 1/1 to: AB032962 from: 1 to: 4595

```
672 AsnAsnProArgAspAspIleAlaThrGluAlaTyrGluAspGluLeuAs 688
|||||
2 ANTAACCCAGAGATGATATGCTACAGAGCATATGAGGATGAGCTAGA 51
|||||
688 pMetGlyArgSerGlySerTyrLeuAsnSerSerIleAsnSerAlaTrps 705
|||||
52 CATGGGCGGATCTGGATCTTACCTGAACAGCAGTATCAATTCAGCCTGGA 101
|||||
705 erGluHisSerLeuAspProGluAspIleArgAspGluLeuLysLeu 721
|||||
102 GTGAGCACAGCTTGGATCCAGAGAGCATTCGGAGCAGAGCTGAATAACTC 151
|||||
722 TyrAlaGlnLeuGluIleTyrLysArgLysLysMetIleThrAsnAsnPr 738
|||||
152 TATGCCCACTGGANATATATAAAGNAAGAGATGATCAACAARACCC 201
|||||
738 oHisLeuGlnLysLysArgCysSerLysLysGlyLeuGlyArgSerIleM 755
|||||
202 CCACCTCCAGAAAAAGCGTGTCTCGAAGAAAGGCGCTAGGTCGTTCATCA 251
|||||
755 etArgArgIleThrGluIleProGluThrValSerArgGlnCysSerLys 771
|||||
252 TGAGACGCATTCGGAGATCCAGAGACAGTCCAGCCGGCAGTCTCTAA 301
|||||
772 GluAspLysGluGlyAlaAspHisGlyThrAlaLysGlyThrAlaLeuIl 788
|||||
302 GAGGACAAGAGGCGCGCCGACCATGTCACAGCCAAAGGCACCTGCCCTCAT 351
|||||
788 eArgLysAsnProProGluSerSerGlyAsnThrGlyLysSerLysGluG 805
|||||
352 CAGGAAGAACCCCCAGAGCTTTCAGGGAACACAGGGAATCCCAAGGAG 401
|||||
805 luThrLeuLysAsnArgValPheSerLeuLysLysSerHisSerThrTyr 821
|||||
402 AGACCTCGAATAACCCAGAGTCTTCTCACTCAAGAAATCCACAGCACTTAT 451
|||||
822 AspHisValArgAspGlnThrGluGluSerSerSerLeuProThrGluSe 838
|||||
452 GACCACGTGAGAGACCAACAGGAGAGTCCAGTAGCCTACCCACAGAAAG 501
|||||
838 rGlnGluGluGluThrThrGluAsnSerThrLeuGluSerLeuSerGlyL 855
|||||
502 CCAAGAGAGGAGAGACAAACAGAAAAATCCACACTGGAATCCCTGTCGGTA 551
|||||
855 ysLysLeuThrGlnLysLeuLysGluAspSerGluAlaGluSerThrGlu 871
|||||
552 AAAAATAACACAAAAAATATAAAGAGAGACGCGAGGCTGAGTCCACGGAG 601
|||||
872 SerValProLeuValCysLysSerAlaSerAlaHisAsnLeuSerSerGI 888
|||||
602 TCGGTCCGCTGTGTGTGCAAGTCAGCAGAGCGCTCAACACTCAGCTCAGA 651
|||||
888 uLysLysThrGlyHisProArgThrSerMetLeuGlnLysSerLeuSerV 905
|||||
652 GAAGAAACTGGGCACCCACCAACATCGATGTTACACAACCTCTCTCAGTG 701
|||||
905 alIleAlaSerAlaLysGluLysThrLeuGlyLeuAlaGlyLysThrGln 921
|||||
702 TCATAGCAAGCGCCAAAGGAGAGAGACTCTTGGATTAGCTGGGAAAAACCAA 751
|||||
922 ThrAlaGlyValGluGluArgThrLysSerGlnLysProLeuProLysAs 938
|||||
752 ACAGCAGGTGTGGAAGACGCACTAATCCAGAAACCTTTGCCAANAAGA 801
|||||
938 pLysGluThrAsnArgAsnHisSerAsnSerAspAsnThrGluThrLysA 955
|||||
802 TAAAGACAAAAACAGAAATCACTCAAAATCTGTATAACACAGAGACTTAA 851
|||||
955 spProAlaProGlnAsnSerAsnProAlaGluProArgLysProGln 971
|||||
```

```
852 ATCCTGCCCCCCCAAACTCAAAATCCTGCGGAGGAGCCAAAGAAAGCCTCAG 901
|||||
972 LysSerGlyIleMetLysGlnArgValAsnProThrThrAlaAsnSe 988
|||||
902 AAATCTGGGATTATGAACAACAAGGGTCAACCCACCACTGCCAATTC 951
|||||
988 rAspLeuAsnProGlyThrThrGlnMetLysAspAsnPheAspIleGly 1005
|||||
952 TGACCTGAACCCAGGACCAACCCAGATGAAGGACAACCTTTGACATTGGGG 1001
|||||
1005 luValCysProThrProGluValTyrAspLeuThrProGlyProValProSer 1021
|||||
1002 AGGTGTCTCTTGGGAGGTTTATGACCTGACCCCTGGTCTGTGCTTCA 1051
|||||
1022 GluSerLysValGlnLysHisValSerIleValAlaSerGluMetGluLy 1038
|||||
1052 GAATCAAAAGTTCAAAAGCACGTTATCTATTGTGGCTTCTGAAAATGGAGAA 1101
|||||
1038 sAsnProThrPheSerSerLeuLysGluLysSerHisHisLysProLysAla 1055
|||||
1102 AAACCCCACTTTTCTCTTAAAGGAGAAATCTCACCAAGCCTAAGGCAG 1151
|||||
1055 laGluValCysGlnGlnSerAsnGlnLysArgIleAspLysAlaGluVal 1071
|||||
1152 CTGAGCTTTGTACAGCAATCCAATCAGAGCGCATAGATAAGGCTGAAGTA 1201
|||||
1072 CysLeuTriPLeuSerGlnGlyGlnSerIleLeuGluAspGluLysLeuLe 1088
|||||
1202 TGCCTTTGGAGAGCAACAGCCAGTCCATTTTGGAGAGATGAGAAAGCTTTT 1251
|||||
1088 uIleSerLysThrProValLeuProGluArgAlaLysGluGluAsnGlyG 1105
|||||
1252 GATTTCCAAGACTCCAGTCTCTCCAGAGAGGCAAAAGAGGAGAACGGAG 1301
|||||
1105 lyGlnProArgAlaAlaAsnValCysAlaGlyClnSerGluGluLeuPro 1121
|||||
1302 GTCAGCTCTGTGAGCAATGTGTGTCTGGGAGAGCGAAGAACTGCCCT 1351
|||||
1122 ProLysAlaValAlaSerLysThrGluAsnGluAsnLeuAsnGlnIleGI 1138
|||||
1352 CCCAAAGCTGTAGCATCAAAACAGAGAATGAAATCTCAACCAATAGG 1401
|||||
1138 yHisGlnGluLysLysThrSerSerSerGluGluAsnValArgGlySerT 1155
|||||
1402 ACACCAAGGAAAAAAGACATCTTCTCTCAGGAGAATGTGCGTGGCTCCT 1451
|||||
1155 yrAsnSerSerAsnAsnPheGlnGlnProLeuThrSerArgAlaGluVal 1171
|||||
1452 ATAACTCAAGTAATAACTTCCAGCAACCTTTAACTACACGACGAGAGTT 1501
|||||
1172 CysProTriPLeuPheGluThrProAlaGlnProAsnAlaGlyArgSerVa 1188
|||||
1502 TGTCTCTGGGAGTTTGAGACCCCAAGCTCAACCAATCTCTGGAGAAGTGT 1551
|||||
1188 laIaLeuProAlaSerSerAlaLeuSerAlaAsnLysIleAlaGlyProA 1205
|||||
1552 AGCTTTACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1601
|||||
1205 rgLysGluGluIleThrAspSerPheLysVal 1215
|||||
1602 GGAAAGAGAGGCTCTGGGATAGTTTTTAAAGTG 1633
```

seq_name: gb_hgt20:AL355587

seq_documentation_block:

LOCUS AL355587 151218 bp DNA
DEFINITION Homo sapiens chromosome 10 clone RP11-561H23, *** SEQUENCING IN
PROGRESS ***, 9 unordered pieces.
ACCESSION AL355587
VERSION AL355587.8 GI:13660940
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens

89130 TCATAATACATGTTTATAACATCTTCTAGGAAAGGTAGGATAGATC 89179
713 713
89180 TGTGGAGCAGAGGAGCGCTGGATTATTTAAATAATCTATGACGACA 89229
713 713
89230 TATCAGCATCTGAAGCAAGCAGATGTTTTAGTAGTCAACGGACTATTAC 89279
713 713
89280 CCTTAAGCCTAATCTAATCATCATGTAACTTTATCAATGGTGCATGA 89329
713 713
89330 TTCTCCATATACTACAGGATATGGTTCACCTATAACAATATGCTGT 89379
713 713
89380 GTTTTCTTCTGGTCCAGGCTCAGTCTGGTCCCTCACAACTTGAATGTACAA 89429
713 713
89430 AGTGCTTGTCTTGTGATGCAAGGTTAGAAATACATGAGTTCTGCTTAAA 89479
713 713
89480 GTCCACACTTTATGTGAACAAACAAACAAAGAAATCAGAAAATACAAATC 89529
713 713
89530 AGAAAGTACAAATCAGAAGGTCTTGCTCCTTGAAGCAAGGTAGAATA 89579
713 713
89580 CGTGAGTTCATGCTTTAAGGTCCACACTTTATGTGAACAAACAAACAAA 89629
713 713
89630 GAATCAGAAAGTACAAATGGTCAGAAAATCGACAAGTACATATGTCAGAG 89679
713 713
89680 AAAGCTGTAGACAGCTTGCCCAAAATTAGAGCCCAAGTTTACTGAATCC 89729
713 713
89730 AATACCATGTAGTATCTTCTATGACTGGACACTTAAATCTTACACTTCT 89779
714 Ile.ArgAspGluLeuLysLysLeuTyrAlaG1 724
89780 TTGAATTGCTTTTGTCTGGCAGCAGCAGCTGAAATACTCTATGCCCA 89829
724 nLeuGluLeuTyrLysArgLysLysMetIleThrAsnAsnProHisLeuG 741
89830 ACTGGAATATATAAAGAAAGCAAGATGATCACAACAAACCCCACTCC 89879
741 nLysLysArgCysSerLysLysGlyLeuGlyArgSerIleMetArgArg 757
89880 AGAAAAAGCGGTGCTCGAAGAGGCGCTAGTCTTCCATCATGAGAGCG 89929
758 IleThrGluIleProGluThrValSerArgGlnCysSerLysGluAspLys 774
89930 ATTACGAGATCCAGACAGACTACGCGGAGTGTCTCTAAGAGGACAA 89979
774 sGluGlyAlaAspHisGlyThrAlaLysGlyThrAlaLeuIleArgLysA 791
89980 CGAGGCGCGCACCATGTCACAGCCAAAGGCACCTGCCCTCATCAGGA 90029
791 snProProGluSerSerGlyAsnThrGlyLysSerLysGluGluThrLeu 807
|||||

90030 ACCCCCCAGAGTCTTTCAGGGAACACAGGGAATCCAAAGGAGGAGCCCTG 90079
808 LysAsnArgValPheSerLeuLysLysSerHisSerThrTyrAspHisVa 824
|||||
90080 AAAACCGAGTCTTCTCACTCAAGAAATCCACAGCAGCTTATGACCAGCT 90129
824 lArgAspGlnThrGluGluSerSerSerLeuProThrGluSerGlnGluG 841
|||||
90130 GAGACACCAACGGAAGAGTCCAGTAGCTTACCCACAGAAACCCAGAGG 90179
841 luGluThrThrGluAsnSerThrLeuGluSerLeuSerGlyLysLysLeu 857
|||||
90180 AGGAGACAAACAGAAATCCACACTGGAATCCCTGTCGGGTAAAAACTA 90229
858 ThrGlnLysLeuLysGluAspSerGluAlaGluSerThrGluSerValPr 874
|||||
90230 ACACAAAAACTAAAGAGACAGCGAGGCTGAGTCCACGGAGTCGGTGCC 90279
874 leuValCysLysSerAlaSerAlaHisAsnLeuSerSerGluLysLysT 891
|||||
90280 GTTGTGTGCAAGTCAGCAAGCGCTCACAACTCAGCTCAGAGAAAGAAA 90329
891 hrGlyHisProArgThrSerMetLeuGlnLysSerLeuSerValIleAla 907
|||||
90330 CTGGCACCCACGAACTCGATGTTACAGAGTCTCTCAGTGTCTATAGCA 90379
908 SerAlaLysGluLysThrLeuGlyLeuAlaGlyLysThrGlnThrAlaG1 924
|||||
90380 ACGGCCAAGGAGAACTCTTGGATTAGCTGGGAAACCCCAACACAGCAGG 90429
924 yValGluGluArgThrLysSerGlnLysProLeuProLysAspLysGluT 941
|||||
90430 TGTGAAGAAGCAACGCACTAAATCCAGAAACCTTTGCCAAAGATAAGAGA 90479
941 hrAsnArgAsnHisSerAsnSerAspAsnThrGluThrLysAspProAla 957
|||||
90480 CAAACAGAAATCACTCAATTTCTGATAACACAGAGACTAAAGATCTCTGC 90529
958 ProGlnAsnSerAsnProAlaGluGluProArgLysProGlnLysSerG1 974
|||||
90530 CCCCAAAATCAATCTCTCGCGAGGAGCCAGAAAGCCTCAGAAATCTGG 90579
974 yIleMetLysGlnGlnArgValAsnProThrThrAlaAsnSerAspLeuA 991
|||||
90580 GATTATGAACAACAAGGGTCAACCCCAACCACTGCCAATTTCTGACCTCA 90629
991 snProGlyThrThrGlnMetLysAspAsnPheAspIleGlyGluValCys 1007
|||||
90630 ACCCAGGCAACCCAGATGAAGGACAACTTTGACATTTGGGAGGTGTGT 90679
1008 ProTtpGluValTyrAspLeuThrProGlyProValProSerGluSerIy 1024
|||||
90680 CCTTGGGAGGTTTATGACCTGACCCCTGTCTGTCTGTCTTCAAGATCAA 90729
1024 sValGlnLysHisValSerIleValAlaSerGluMetGluLysAsnProT 1041
|||||
90730 AGTTCAAAAGCACGTAATCTATTGTGGCTTCTGAAATGGAGAAAAACCCCA 90779
1041 hrPheSerLeuLysGluLysSerHisLysProLysAlaAlaGluVal 1057
|||||
90780 CTTTTCCTTAAAGGAGAAATCTCACCAAGCCTTAAGGCAGCTGAGGTT 90829
1058 CysGlnGlnSerAsnGlnLysArgIleAspLysAlaGluValCysLeuTr 1074
|||||
90830 TGTACGCAATCCAATCAGACGCGCATAGATAAGGCTGAAGTATGCTTTG 90879
1074 pGluSerGlnGlyGlnSerIleLeuGluAspGluLysLeuLeuIleSerL 1091
|||||
90880 CGAGAGCCAAAGGCCAGTCCATTTTGGAGAGATGAGAAAGCTTTTGTATTCCA 90929
1091 ysThrProValLeuProGluArgAlaLysGluGluAsnGlyGlyGlnPro 1107
|||||
90930 AGACTCCAGTTCTCCAGAGAGGCGCAAGAGAGAGAACGAGGAGTACGCT 90979

```
1108 ArgAlaAlaAsnValCysAlaGlyClnSerGluGluLeuProProLysAl 1124
|||||
90980 CGTGCAGCAATGTGTGCTGGCAGAGCGAAGAACTGCCCCCAAGC 91029
|||||
1124 aValAlaSerLysThrGluAsnGluAsnLeuAsnGlnIleGlyHisGlnG 1141
|||||
91030 TGTAGCATCAAAACAGAGATGAATCTCAACCAATAGACACACAGG 91079
|||||
1141 LuLysLysThrSerSerSerGluGluAsnValArgGlySerTyrAsnSer 1157
|||||
91080 AAAAAAGACATCTTCTCTCAGGAGAAATGTGCGTGGCTCTATAACTCA 91129
|||||
1158 SerAsnAsnPheGlnGlnProLeuThrSerArgAlaGluValCysProTr 1174
|||||
91130 AGTAATAACTTCCAGACACCTTTAATCATCAGCAGCAGAGGTTTGTCTTG 91179
|||||
1174 pGluPheGluThrProAlaGlnProAsnAlaGlyArgSerValAlaLeuP 1191
|||||
91180 GGAGTTTGAGACCCAGCCAGCTCAACCAATGCTGGGAAGATGATGCTTTAC 91229
|||||
1191 roAlaSerSerAlaLeuSerAlaAsnLysIleAlaGlyProArgLysGlu 1207
|||||
91230 CTGCCTCTTCTGCTTAAGTGCATAATAGATAGCAGGCGCTAGGAAGAA 91279
|||||
1208 GluLeuTrpAspSerPheLysVal 1215
|||||
91280 GAGATCTGGGATAGTTTAAAGTG 91303
|||||
seq_name: gb_htg20:AL355542

seq_documentation_block:
LOCUS AL355542 185146 bp DNA HTG 13-APR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-257J14, *** SEQUENCING IN
PROGRESS ***, 19 unordered pieces.
ACCESSION AL355542
VERSION AL355542.9 GI:13624986
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 185146)
Plumb,B.
Direct Submission
Submitted (12-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 14, 2001 this sequence version replaced gi:13624506.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA257J14
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 175310 bases at least Q40
Consensus quality: 178563 bases at least Q30
Consensus quality: 180883 bases at least Q20
Insert size: 183346; sum-of-contigs
Insert size: 145596; 18.0% error; agarose-fp
Quality coverage: 4.36x in Q20 bases; sum-of-contigs Quality
coverage: 7.64x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
```

```
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 11771: contig of 11771 bp in length
11772 11871: gap of 100 bp
11872 61631: contig of 49760 bp in length
61632 61731: gap of 100 bp
61732 79327: contig of 17596 bp in length
79328 79427: gap of 100 bp
79428 83309: contig of 3882 bp in length
83310 83409: gap of 100 bp
83410 109528: contig of 26119 bp in length
109529 109628: gap of 100 bp
109629 112808: contig of 3180 bp in length
112809 112908: gap of 100 bp
112909 118905: contig of 5997 bp in length
118906 119005: gap of 100 bp
119006 133743: contig of 14738 bp in length
133744 133843: gap of 100 bp
133844 143213: contig of 9370 bp in length
143214 143313: gap of 100 bp
143314 158185: contig of 14872 bp in length
158186 158285: gap of 100 bp
158286 162061: contig of 3776 bp in length
162062 162161: gap of 100 bp
162162 164747: contig of 2586 bp in length
164748 164847: gap of 100 bp
164848 167658: contig of 2811 bp in length
167659 167758: gap of 100 bp
167759 170009: contig of 2251 bp in length
170010 170109: gap of 100 bp
170110 173787: contig of 3678 bp in length
173788 173887: gap of 100 bp
173888 176350: contig of 2463 bp in length
176351 176450: gap of 100 bp
176451 179441: contig of 2991 bp in length
179442 179541: gap of 100 bp
179542 182272: contig of 2731 bp in length
182273 182372: gap of 100 bp
182373 185146: contig of 2774 bp in length.

FEATURES
            Location/Qualifiers
             source          1..185146
                             /organism="Homo sapiens"
                             /db_xref="taxon:9606"
                             /chromosome="10"
                             /clone="RP11-257J14"
                             /clone_lib="RPCI-11.1"
             misc_feature    1..11771
                             /note="assembly_fragment:00155
                             fragment_chain:1"
                             11872..61631
                             /note="assembly_fragment:00988
                             fragment_chain:1"
                             61732..79327
                             /note="assembly_fragment:01331
                             fragment_chain:1"
                             79428..83309
                             /note="assembly_fragment:03102
                             fragment_chain:1"
                             83410..109528
                             /note="assembly_fragment:02919
                             fragment_chain:1"
                             109629..112808
                             /note="assembly_fragment:01757
                             fragment_chain:1"
                             112909..118905
                             /note="assembly_fragment:01162
                             fragment_chain:1"
                             119006..113743
                             /note="assembly_fragment:01246
                             fragment_chain:1"
                             133844..143213
                             /note="assembly_fragment:01266
                             fragment_chain:1"
```

```
misc_feature      fragment_chain:1"
143314. .158185
/note="assembly_fragment:03087
fragment_chain:1"
158286. .162061
/note="assembly_fragment:00095"
162162. .164747
/note="assembly_fragment:00400"
164848. .167658
/note="assembly_fragment:00509"
167759. .170009
/note="assembly_fragment:01477"
170110. .173787
/note="assembly_fragment:01903"
173888. .176350
/note="assembly_fragment:02102"
176451. .179441
/note="assembly_fragment:02396"
179542. .182272
/note="assembly_fragment:02514"
182373. .185146
/note="assembly_fragment:03322"
BASE COUNT 50876 a 38769 c 39206 g 54480 t 1815 others
ORIGIN

alignment_scores:
  Quality: 2435.00      Length: 502
  Ratio: 4.909          Caps: 1
Percent Similarity: 98.805 Percent Identity: 97.410

alignment_block:
US-09-775-181-2 x AL355542/rev ..
Align seg 1/1 to reverse of: AL355542 from: 1 to: 185146

715 ArgAspGluLeuLysLysLeuTyrAlaGlnLeuGluIleTyrLysArgLy 731
:::
164477 CAGGACGAGTGAAGAACTCTATGCCCCAATCGAAATATATAAAGAAA 164428

731 sLysMetIleThrAsnAsnProHisLeuGlnLysLysArgCysSerLysL 748
|||||
164427 GAGATGATCACAAACACCCCCACCTCCAGAAAAGCGGTGCTCGAAGA 164378

748 yGlyLeuGlyArgSerIleMetArgArgIleThrGluIleProGluThr 764
|||||
164377 AGGCCCTAGTCTGCTCATCATGAGCGCATTAACGAGATCCCGAGACA 164328

765 valSerArgGlnCysSerLysGluAspLysGluGlyAlaAspHisGlyTh 781
|||||
164327 GTCCAGCCGCGAGTCTCTAAAGAGGACAAAGGAGGCGCCGACCATGC 164278

781 rAlaLysGlyThrAlaLeuIleArgLysAsnProProGluSerSerGlyA 798
|||||
164277 AGCCAAAGGACTGCCCTCATCAGGAAGAACCCCCCAGAGTCTTCAGGA 164228

798 snThrGlyLysSerLysGluGluThrLeuLysAsnArgValPheSerLeu 814
|||||
164227 ACACAGGAAATCCAGGAGGAGACCCCTGAAAACCGAGTCTCTCACTC 164178

815 LysLysSerHisSerThrTyrAspHisValArgAspGlnThrGluGluSe 831
|||||
164177 AGAAATCCACAGCAGCACTTATGACCACTGAGAGACCAACGGAAGATC 164128

831 rSerSerLeuProThrGluSerGlnGluGluThrThrGluAsnSerT 848
|||||
164127 CAGTACCTTACCACAGAAGCCAGAGGAGGAGACACAGAAAATTCCTCA 164078

848 hrLeuGluSerLeuSerGlyLysLysLeuThrGlnLysLeuLysGluAsp 864
|||||
164077 CACTGGAATCCCTGTCGGCTAAAAAATAACACAAAAAATAAAAGAGAC 164028

865 SerGluAlaGluSerThrGluSerValProLeuValCysLysSerAlase 881
|||||
```

```
|||||
164027 AGCGAGCTCAGTCCACGAGTCGGTCCGTTGGTGTGCAAGTCAGCAAG 163978

881 rAlaHisAsnLeuSerSerCylulysLysThrGlyHisProArgThrSerM 898
|||||
163977 CGCTCACACCTCAGCTCAGAGAAGAAACTGGGCACCCACGAAATCGA 163928

898 etLeuGlnLysSerLeuSerValIleAlaSerAlaLysGluLysThrLeu 914
|||||
163927 TGTACAGAAGTCTCTCAGTGTCTATAGCAAGCCCAAGGAGAGACTCTT 163878

915 GlyLeuAlaGlyLysThrGlnThrAlaGlyValGluGluArgThrLysSe 931
|||||
163877 GSATTAGCTGGGAAAAACCCACACAGCAGGTGTGGAAGACGCACATAATC 163828

931 rGlnLysProLeuProLysAspLysGluThrAsnArgAsnHisSerAsnS 948
|||||
163827 CCAGAAACCTTTGCCAAAAGATAAAGAGACAAACAGAAATCACTCAAAAT 163778

948 erAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsnProAla 964
|||||
163777 CTGATAACACAGAGACTAAGATCCTGCCCCCAAACTCAAATCCTGCG 163728

965 GluGluProArgLysProGlnLysSerGlyIleMetLysGlnGlnArgVa 981
|||||
163727 GAGGAGCCAAAGAGCCTCAGAAATCTGGGATTATGAAACAACAAAGGTC 163679

981 lAsnProThrThrAlaAsnSerAspLeuAsnProGlyThrThrGlnMetL 998
|||||
163678 .AACCCACCACTGCCAATTTCTGACCTGAACCCAGCCACCCACCATGA 163629

998 ysAspAsnPheAspIleGlyGluValCysProTyrGluValTyrAspLeu 1014
|||||
163628 AGGACAACCTTTGACATTGGGAGGTGTCTCTTGGAGGTTTATGACCTG 163579

1015 ThrProGlyProValProSerGluSerLysValGlnLysHisValSerIl 1031
|||||
163578 ACCCTTGGTCTGTGCTTTCAGAAATCAAAAGTTCAAAAGCAGCATCTAT 163529

1031 eValAlaSerGluMetGluLysAsnProThrPheSerLeuLysGluLysS 1048
|||||
163528 TGTGGCTTCTGAATYGGAGAAAACCCCACTTTTCTTAAAGGAGAAAT 163479

1048 erHisLysProLysAlaAlaGluValCysGlnGlnSerAsnGlnLys 1064
|||||
163478 CTCAACCAAGCCTAAGGAGGTGATGTTTGTTCAGCAATCAATCAATAG 163429

1065 ArgIleAspLysAlaGluValCysLeuTyrGluSerGln. GlyGlnSerI 1081
|||||
163428 CGCATAGATAAGGCTGAAGTATGCCTTTGGGAGAGCCAAAGGCCAGTCCA 163379

1081 leLeuGluAspGluLysLeuLeuIleSerLysThrProValLeuProGlu 1097
|||||
163378 TTTGGGAAGATGAGAAGCTTTTGAATTTCCAGACTCCAGTTCTCCAGAG 163329

1098 ArgAlaLysGluGluAsnGlyGlyGlnProArgAlaAlaAsnValCysAl 1114
|||||
163328 .AGGCAAAAGAGAGAGACGAGAGTCAGCTCTGTCAGCAATGTGTGTC 163279

1114 aglyGlnSerGluGluLeuProProLysAlaValAlaSerLysThrGluA 1131
|||||
163278 TGGCAG .AGCGAAGAACTGCCCCCAAGCTGTAGCATCAAAAACAGAGA 163230

1131 snGluAsnLeuAsnGlnIleGlyHisGlnGluLysLysThrSerSerSer 1147
|||||
163229 ATGAAAATCTCAACCAATAGGACACCCAGGAAAAAAGACATCTTCTCT 163180

1148 GluGluAsnValArgGlySerTyrAsnSerSerAsnAsnPheGlnGlnPr 1164
|||||
163179 GAGGAGAAATGTCGGTGGCTCTTATAACTCAAGTAATAACTTCCAGCAACC 163130

1164 oLeuThrSerArgAlaGluValCysProTyrGluPheGluThrProAlaG 1181
|||||
```


653 AAGAGGAATCTCACCACAGCCTAAGCAGCTGAAGTTTGTGACCAAT 702
1061 erAsnGlnLysArgIleAspLysAlaGluValCysLeuTrpGluSerGln 1077
|||||
703 CCAATCAGAGTCCACAGATAAGGCTGAAGTAGCTTGGGGGATCCAA 752
1078 GlyClnSerIleLeuGluaspGluLysLeuLeuIleSerLysThrProVa 1094
|||||
753 GGCCAGTCCATTGGAAGATGAGAGCATTTTCATTTCGAAGACTCCAGT 802
1094 lLeuProGluArgAlaLysClnGluAsnGlyGlnProArgAlaAla 1111
|||||
803 TCTCCAGAGAGGGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 952
1111 snValCysAlaGlyGlnSerGluGluLeuProProLysAlaValAlaSer 1127
:|||||
853 AAGTGTGTCTGGGCAGAGTGAAGAACTGCCCCCAAGCTGTAGCATCA 902
1128 LysThrGluAsnGluAsnLeuAsnGlnIleGlyHisGlnGluLysLysTh 1144
|||||
903 AAACACAGAGAATGAAATCTCAACCAATAGGACACAGCAAAAAAGAC 952
1144 rSerSerSerGluGluAsnValArgGlySerTyrAsnSerSerAsnAsp 1161
|||||
953 A...TCTTTGAGGAGAAATGCGTGCTCTCTATACTCAAGTAATACT 999
1161 heGlnGlnProLeuThrSerArgAlaGluValCysProTrpGluPheGlu 1177
|||||
1000 TCCAGCAACCTTTAATGTCCAGCAGCAGAGGTGTGCCCTTGGGAGTTTGAG 1049
1178 ThrProAlaGlnProAsnAlaGlyArgSerValAlaLeuProAlaSerSe 1194
|||||
1050 ACCCCAGCTCAACCAATCTCGAAGAGTGTACCTTACCTGTTTCTTC 1099
1194 rAlaLeuSerAlaAsnLysIleAlaGlyProArgLysGluLleTrpA 1211
|||||
1100 TGCTCTAAGTCAAGTAGATAGCAGGCGCTCGGAAGAAGAAGTCTGGG 1149
1211 spSerPheLysVal 1215
|||||
1150 ATACTTTTAAAGTG 1163

seq_name: gb_hgt21:ALJ358932

seq_documentation_block:
LOCUS ALJ358932 111117 bp DNA HTG 23-JAN-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-395P8, *** SEQUENCING IN
PROGRESS ***, 25 unordered pieces.
VERSION ALJ358932
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 111117)
Direct Submission
Sims,S.
Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk clone
requests: clonerequest@sanger.ac.uk
On Aug 27, 2000 this sequence version replaced gi:9926667.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA395P8
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 90261 bases at least Q40
Consensus quality: 99226 bases at least Q30
Consensus quality: 10490 bases at least Q20
Insert size: 108717; sum-of-contigs
Insert size: 178772; 24.0% error; agarose-fp
Quality coverage: 1.84x in Q20 bases; sum-of-contigs Quality
coverage: 1.42x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 2287: contig of 2287 bp in length
* 2288 2387: gap of 100 bp
* 2388 6317: contig of 3930 bp in length
* 6318 6417: gap of 100 bp
* 6418 9725: contig of 3308 bp in length
* 9726 9825: gap of 100 bp
* 9826 16607: contig of 6782 bp in length
* 16608 16707: gap of 100 bp
* 16708 20159: contig of 3452 bp in length
* 20160 20259: gap of 100 bp
* 20260 23641: contig of 3382 bp in length
* 23642 23741: gap of 100 bp
* 23742 27395: contig of 3658 bp in length
* 27400 27499: gap of 100 bp
* 27500 32754: contig of 5255 bp in length
* 32755 32854: gap of 100 bp
* 32855 35240: contig of 2386 bp in length
* 35241 35340: gap of 100 bp
* 35341 40948: contig of 5608 bp in length
* 40949 41048: gap of 100 bp
* 41049 43758: contig of 2710 bp in length
* 43759 43858: gap of 100 bp
* 43859 48684: contig of 4826 bp in length
* 48685 48784: gap of 100 bp
* 48785 52405: contig of 3621 bp in length
* 52406 52505: gap of 100 bp
* 52506 56282: contig of 3777 bp in length
* 56283 56382: gap of 100 bp
* 56383 59669: contig of 3287 bp in length
* 59670 59769: gap of 100 bp
* 59770 63200: contig of 3431 bp in length
* 63201 63300: gap of 100 bp
* 63301 68131: contig of 4831 bp in length
* 68132 68231: gap of 100 bp
* 68232 70678: contig of 2447 bp in length
* 70679 70778: gap of 100 bp
* 70779 72875: contig of 2097 bp in length
* 72876 72975: gap of 100 bp
* 72976 79043: contig of 6068 bp in length
* 79044 79143: gap of 100 bp
* 79144 88143: contig of 9000 bp in length
* 88144 88243: gap of 100 bp
* 88244 91654: contig of 3411 bp in length
* 91655 91754: gap of 100 bp
* 91755 102397: contig of 10643 bp in length
* 102398 102497: gap of 100 bp
* 102498 106184: contig of 3687 bp in length
* 106185 106284: gap of 100 bp
* 106285 111117: contig of 4833 bp in length.

FEATURES
Location/Qualifiers
1..11117
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-395P8"
/clone_lib="RPC1-11.2"
1..2287
misc_feature

```
/note="assembly_fragment:00042
fragment_chain:1"
misc_feature 2388..6317
/note="assembly_fragment:00050
fragment_chain:1"
misc_feature 6418..9725
/note="assembly_fragment:00150
fragment_chain:2"
misc_feature 9826..16607
/note="assembly_fragment:00506
fragment_chain:2"
misc_feature 16708..20159
/note="assembly_fragment:00247
fragment_chain:3"
misc_feature 20260..23641
/note="assembly_fragment:00502
fragment_chain:3"
misc_feature 23742..27399
/note="assembly_fragment:00625
fragment_chain:4"
misc_feature 27500..32754
/note="assembly_fragment:00510
fragment_chain:4"
misc_feature 32855..35240
/note="assembly_fragment:00035"
misc_feature 35341..40948
/note="assembly_fragment:00100"
misc_feature 41049..43758
/note="assembly_fragment:00117"
misc_feature 43859..48684
/note="assembly_fragment:00200"
misc_feature 48785..52405
/note="assembly_fragment:00227"
misc_feature 52506..56282
/note="assembly_fragment:00235"
misc_feature 56383..59669
/note="assembly_fragment:00253"
misc_feature 59770..63200
/note="assembly_fragment:00303"
misc_feature 63301..68131
/note="assembly_fragment:00304"
misc_feature 68232..70678
/note="assembly_fragment:00421"
misc_feature 70779..72875
/note="assembly_fragment:00444"
misc_feature 72976..79043
/note="assembly_fragment:00512"
misc_feature 79144..88143
/note="assembly_fragment:00518"
misc_feature 88244..91654
/note="assembly_fragment:00562"
misc_feature 91755..102397
/note="assembly_fragment:00573"
misc_feature 102498..106184
/note="assembly_fragment:00584"
misc_feature 106285..111117
/note="assembly_fragment:00600"
BASE COUNT 32506 a 20474 c 20158 g 35546 t 2433 others
ORIGIN

alignment_scores:
  Quality: 1607.00      Length: 301
  Ratio: 5.339          Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-775-181-2 x AL358932 ..
|||||
Align seg 1/1 to: AL358932 from: 1 to: 111117

1 MetGlyAlaMetAlaTyrProLeuLeuLeuLeuLeuAlaGlnLe 17
|||||
```

```
83403 ATGGAGCCCATGGCTTACCCCTTACTCCTCTGCTCTCTGCTGCTCAGCT 83452
17 uGlyLeuGlyAlaValGlyAlaSerArgAspProGlnGlyArgProAsps 34
|||||
83453 GGGATTGGAGCTGTTGGCGCCAGCGCCACCCCAAGGACGGCGGATT 83502
34 erProArgGluArgThrProLysGlyLysProHisAlaGlnGlnProGly 50
|||||
83503 CCCCTCGAGAGAGACCCGGAAGGGGAAGCGCACGCCACGACGCGGGT 83552
51 ArgAlaSerAlaSerAspSerSerAlaProTrpSerArgSerThrAspG1 67
|||||
83553 CGACCTCTGCCTCGACTCCTCGCTCCCTGGAGCCGCTCCACCGATGG 83602
67 yThrIleLeuAlaGlnLysLeuAlaGluValProMetAspValAlas 84
|||||
83603 CACCATCTTGGCGCAGAAACTCGCCGAGGAGTGCCTCATGGACGTGGCT 83652
84 erTyrLeuTyrThrGlyAspSerHisGlnLeuLysArgAlaAsnCySer 100
|||||
83653 CTTACCTCTACACCGGGGACTCCCAACGAGCTGAAGCGAGCAACTGTCC 83702
101 GlyArgTyrGluLeuAlaGlyLeuProGlyLysTrpProAlaLeuAlaSe 117
|||||
83703 GCCCGCTACGAGTTGGCGGGCTGCCGGGAAGTGGCCAGCCCTGGCCAG 83752
117 rAlaHisProSerLeuHisArgAlaLeuAspThrLeuThrHisAlaThra 134
|||||
83753 CGCGCACCCCTCCTTGCACCGGGCGCTGGACACACTGACACACGCCACCA 83802
134 snPheLeuAsnValMetLeuGlnSerAsnLysSerArgGluGlnAsnLeu 150
|||||
83803 ACTTCCTCACGTGATGTCAGAGCAATAAGTCGCGGAGCAACTTG 83852
151 GlnAspAspLeuAspTyrTrpGlnAlaLeuValTrpSerLeuLeuGluG1 167
|||||
83853 CAGGACGACCTGGATTGGTACCAGGCGCTGGTGTGGAGCCCTCTGGAGGG 83902
167 yGluProSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeus 184
|||||
83903 CGAGCCCAAGCATCTCCCGGGGGCCATCACCTTCAGCACCGATTGCGTGT 83952
184 erAlaProAlaProGlnValPheLeuGlnAlaThrArgGluGluSerArg 200
|||||
83953 CCGCACCGGGCCCAAGGCTCTCTCCAGGCCACGCGCGAGGAGCGCG 84002
201 IleLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaTh 217
|||||
84003 ATCCTGCTCCCAAGACCTCTCTCTCCGCCACCCACCTTGGCCAAAGCCAC 84052
217 rLeuGluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisL 234
|||||
84053 TCTGGAGACCGAGTGGTTCACAGGCTCCCGGCGCAAGTGGAGGCCCACT 84102
234 euHisArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrp 250
|||||
84103 TACACCGCGCGGCCCAATAACAGGGGCCCGGGGCTGGGCCACAGCTGG 84152
251 ArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSerPr 267
|||||
84153 CGGCGCAAGGACGGGCTCGCGGGGACAAGAGCCACTTCAAGTGGTGTCC 84202
267 oProTyrLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeuValT 284
|||||
84203 GCCTTATCTGGAGTGGAGAACGGGAGTTACAAGCCCGGGTGGCTGGTTA 84252
284 hrLeuSerSerAlaIleTyrGlyLeuGlnProAsnLeuValProGluPhe 300
|||||
84253 CTCCTTCTCTGCCATCTACGGGTTGACGCTTAACCTGGTCCCGGAATTC 84302
301 Arg 301
||||
84303 AGG 84305
```

```
seq_name: gb_htg20:AL139821
seq_documentation_block:
LOCUS AL139821 199526 bp DNA HTG 17-MAR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-80K21, *** SEQUENCING IN
PROGRESS ***, 6 unordered pieces.
ACCESSION AL139821
VERSION AL139821.7 GI:13396382
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Plumb.B.
Direct Submission
Submitted (15-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
On Mar 20, 2001 this sequence version replaced gi:10185474.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA80K21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 197300 bases at least Q40
Consensus quality: 198299 bases at least Q30
Consensus quality: 198676 bases at least Q20
Insert size: 199026; sum-of-contigs
Insert size: 14047; 11.6% error; agarose-fp
Quality coverage: 7.42x in Q20 bases; sum-of-contigs Quality
coverage: 10.32x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 37012: contig of 37012 bp in length
* 37013 37112: gap of 100 bp
* 37113 163560: contig of 126448 bp in length
* 163561 163660: gap of 100 bp
* 163661 171563: contig of 7903 bp in length
* 171564 171663: gap of 100 bp
* 171664 186786: contig of 15123 bp in length
* 186787 186886: gap of 100 bp
* 186887 196057: contig of 9171 bp in length.
* 196058 196157: gap of 100 bp
* 196158 199526: contig of 3369 bp in length.
*
* Location/Qualifiers
  1. .199526
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="10"
    /clone="RP11-80K21"
    /clone_lib="RPC1-11.1"
    /note="assembly_fragment:00668"
  misc_feature
    1. .37012
      fragment_chain:1
      clone_end:T7
      vector_side:left
    37113 - 163560
      /note="assembly_fragment:02214"
  misc_feature
```

```
fragment_chain:1
clone_end:SP6
vector_side:right"
163661..171563
/note="assembly_fragment:00510"
171664..186786
/note="assembly_fragment:01538"
186887..196057
/note="assembly_fragment:01931"
196158..199526
/note="assembly_fragment:03255"
BASE COUNT 57428 a 38898 c 39349 g 63347 t 504 others
ORIGIN
alignment_scores:
  Quality: 1607.00      Length: 301
  Ratio: 5.339          Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-775-181-2 x AL139821
Align_seg 1/1 to: AL139821 from: 1 to: 199526
1 MetGlyAlaMetAlaTyrProLeuLeuLeuCysLeuLeuAlaGlnLe 17
|||||
97158 ATGGGAGCCATGGCTTACCCCTTACTCCTCTGCTCTGCTCAGCT 97207
17 uGlyLeuGlyAlaValGlyAlaSerArgAspProGlnGlyArgProAsps 34
|||||
97208 GGGATTGGGAGCTGTGGCGCGAGCGCGCCCAAGGACGCCGGATT 97257
34 exProArgGluArgThrProLysGlyLysProHisAlaGlnGlnProGly 50
|||||
97258 CCCCTCGAGAGAGAGCCCGAAGGGAGAGCCGACGCCGAGCGCGGT 97307
51 ArgAlaSerAlaSerAspSerSerAlaProTrpSerArgSerThrAspG1 67
|||||
97308 CGAGCTCTCGCTCGGACTCCTCGGCTCCTCGAGCGCTCCACCGATGG 97357
67 yThrIleLeuAlaGlnLysLeuAlaGluValProMetAspValAlaAs 84
|||||
97358 CACCATCTTGGCGCAGAACTCGCGAGGAGGTGCCCATGGAGCTGGCCT 97407
84 exTyrLeuTyrThrGlyAspSerHisGlnLeuLysArgAlaAsnCysser 100
|||||
97408 CTTACTCTACACGGGGACTCCACAGCTGAGCGAGCCCACTGCTCC 97457
101 GlyArgTyrGluLeuAlaGlyLeuProGlyLysTrpProAlaLeuAlaSe 117
|||||
97458 GCCCGCTACGAGTTGGCGGCGCTGCCGGGAAGTGCCAGCCCTGGCCAG 97507
117 rAlaHisProSerLeuHisArgAlaLeuAspThrLeuThrHisAlaThrA 134
|||||
97508 CGCGCACCCCTCTCTTGCACCGGGCGCTGGACACACTGACACACGCCACA 97557
134 snPheLeuAsnValMetLeuGlnSerAsnLysSerArgGluGlnAsnLeu 150
|||||
97558 ACTTCTCAACCTGATGCTGCAGACAATAAGTCGGGAGCAGAACTTG 97607
151 GlnAspAspLeuAspTrpTyrGlnAlaLeuValTrpSerLeuLeuGlu1 167
|||||
97608 CAGGACGACCTGGATGGTACCAGCGCGTGGTGTGGAGCCCTTCTGGAGG 97657
167 yGluProSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeuS 184
|||||
97658 CGAGCCGACGATCTCCCGGGCGGCATCACCTTACGACCCGATTCGCTGT 97707
184 exAlaProAlaProGlnValPheLeuGlnAlaThrArgGluSerArg 200
|||||
97708 CCGCACCGGGCCCCACAGGTCTTCTCCAGGCCACCGCGAGGAGAGCGCG 97757
```

```
201 lleLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaTh 217
97758 ATCTGTCTCAAGACCTCTCTCTCCGACCCACCTGGCCAAAGCCAC 97807

217 rLeuGlnThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisL 234
97808 TCTGGAGACCGAGTGTTCACGGCTCCGGCGGAAGTGGAGGCCCACT 97857

234 euHisArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrp 250
97858 TACACCGCGCGGCCCAATCAGGGGCCCGGGCTGGGCCACAGCTGG 97907

251 ArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSerPr 267
97908 CGGCGCAAGGACGGCTCGCGGGGGACAAAGACCACTTCAAGTGGTCTCC 97957

267 oProTrpLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeuValT 284
97958 GCCTTATCTGAGTGGAGTGGAGACAGGGAGTTCACAGCCCGGGTGGCTGTTA 98007

284 hrLeuSerSerAlaIleTyrGlyLeuGlnProAsnLeuValProGluPhe 300
98008 CTCCTTCTCTGCCATCTACGGGTTCAGCGCTAACCTGGTCCCGGAATTC 98057

301 Arg 301
|||
- 98058 AGG 98060

seq_name: gb_htg21:AL359974

seq_documentation_block:
LOCUS AL359974 143878 bp DNA HTG 23-JAN-2001
DEFINITION Homo sapiens chromosome 10 clone RP13-236A4, *** SEQUENCING IN
PROGRESS ***, 25 unordered pieces.
ACCESSION AL359974
VERSION AL359974.4 GI:9864577
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 143878)
Sims,S.
Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 22, 2000 this sequence version replaced gi:9795088.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BB236A4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 130224 bases at least Q40
Consensus quality: 135875 bases at least Q30
Consensus quality: 138890 bases at least Q20
Insert size: 141478; sum-of-contigs
Insert size: 205930; 23.5% error; agarose-fp
Quality coverage: 3.17x in Q20 bases; sum-of-contigs Quality
coverage: 2.50x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
```

```
* as soon as it is available and the accession number will
* be preserved.
1 12130: contig of 12130 bp in length
12131 12230: gap of 100 bp
12231 20274: contig of 8044 bp in length
20275 20374: gap of 100 bp
20375 23062: contig of 2688 bp in length
23063 23162: gap of 100 bp
23163 26227: contig of 3065 bp in length
26228 26327: gap of 100 bp
26328 33260: contig of 6933 bp in length
33261 33360: gap of 100 bp
33361 35571: contig of 2211 bp in length
35572 35671: gap of 100 bp
35672 38834: contig of 3163 bp in length
38835 38934: gap of 100 bp
38935 44371: contig of 5437 bp in length
44372 44471: gap of 100 bp
44472 47729: contig of 3258 bp in length
47730 47829: gap of 100 bp
47830 52657: contig of 4828 bp in length
52658 52757: gap of 100 bp
52758 54776: contig of 2019 bp in length
54777 54876: gap of 100 bp
54877 57327: contig of 2451 bp in length
57328 57427: gap of 100 bp
57428 60599: contig of 3172 bp in length
60600 60699: gap of 100 bp
60700 64861: contig of 4162 bp in length
64862 64961: gap of 100 bp
64962 71930: contig of 6969 bp in length
71931 72030: gap of 100 bp
72031 75429: contig of 3399 bp in length
75430 75529: gap of 100 bp
75530 78541: contig of 3012 bp in length
78542 78641: gap of 100 bp
78642 80961: contig of 2320 bp in length
80962 81061: gap of 100 bp
81062 83099: contig of 2038 bp in length
83100 83199: gap of 100 bp
83200 95484: contig of 12285 bp in length
95485 95584: gap of 100 bp
95585 97735: contig of 2151 bp in length
97736 97835: gap of 100 bp
97836 102251: contig of 4416 bp in length
102252 102351: gap of 100 bp
102352 131392: contig of 29041 bp in length
131393 131492: gap of 100 bp
131493 137581: contig of 6089 bp in length
137582 137681: gap of 100 bp
137682 143878: contig of 6197 bp in length.

FEATURES
             Location/Qualifiers
             1..143878
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="10"
             /clone_lib="RP13-236A4"
             /clone_lib="RP13-13.1"
             1..12130
             /note="assembly fragment:01047
             fragment_chain:1"
             12231..20274
             /note="assembly fragment:00007
             fragment_chain:1"
             20375..23062
             /note="assembly fragment:01237
             fragment_chain:1"
             23163..26227
             /note="assembly fragment:01103
             fragment_chain:1"
             26328..33260
             /note="assembly fragment:00266
             fragment_chain:1"
```

```
misc_feature 33361..35571
/note="assembly_fragment:01006
fragment_chain:1"
misc_feature 35672..38834
/note="assembly_fragment:00201
fragment_chain:2"
misc_feature 38935..44371
/note="assembly_fragment:01057
fragment_chain:2"
misc_feature 44472..47729
/note="assembly_fragment:00179
fragment_chain:2"
misc_feature 47830..52657
/note="assembly_fragment:00444
fragment_chain:3"
misc_feature 52758..54776
/note="assembly_fragment:00414
fragment_chain:3"
misc_feature 54877..57327
/note="assembly_fragment:00586
fragment_chain:4"
misc_feature 57428..60599
/note="assembly_fragment:00641
fragment_chain:4"
misc_feature 60700..64861
/note="assembly_fragment:00717
fragment_chain:5"
misc_feature 64962..71930
/note="assembly_fragment:01116
fragment_chain:5"
misc_feature 72031..75429
/note="assembly_fragment:00156"
fragment_chain:5"
misc_feature 75530..78541
/note="assembly_fragment:00233"
fragment_chain:5"
misc_feature 78642..80961
/note="assembly_fragment:00273"
fragment_chain:5"
misc_feature 81062..83099
/note="assembly_fragment:00279"
fragment_chain:5"
misc_feature 83200..95484
/note="assembly_fragment:00515"
fragment_chain:5"
misc_feature 95585..97735
/note="assembly_fragment:00654"
fragment_chain:5"
misc_feature 97836..102251
/note="assembly_fragment:00675"
fragment_chain:5"
misc_feature 102352..131392
/note="assembly_fragment:00719"
fragment_chain:5"
misc_feature 131493..137561
/note="assembly_fragment:00767"
fragment_chain:5"
misc_feature 137682..143878
/note="assembly_fragment:00852
clone_end:SP6
vector_side:right"
BASE COUNT 44271 a 25988 c 25788 g 45414 t 2417 others
ORIGIN

alignment_scores:
  Quality: 413.50      Length: 447
  Ratio: 1.854        Caps: 19
  Percent Similarity: 49.888      Percent Identity: 30.872

alignment_block:
US-09-775-181-2 x AL359974/rev ..
Align seg 1/1 to reverse of: AL359974 from: 1 to: 143878
372 ArqGlyProAspGlnHisThrLysSerThrLysAspValSerGluCl 388
89283 AGGGGTCGGGATCAGCATATTTTCCAGGAAGTACAAAAGATGTGTGACAAGA 89234
388 uAlaTrpValCysLeuProCysArgGluGlyCysProPheCysAlaAspA 405
89233 AGCCATATGTCGCTACCTTTCAGGAGGAGGCTGCCCTCTCTGCTGATG 89184
```

```
405 spSerProCysPheValGlnGluAspLysTyrLeuArgLeuAlaIleIle 421
|||||
89183 ACAGCCCATGCTTCGTCAGGAAGATAAGTATTATGACTTGCCATCATC 89134
422 SerPheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeuValVa 438
|||||
89133 TCCTTCCAAGGCTGTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 89084
438 lTyrHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuLeuL 455
|||||
89083 CTACCACTTTCCGAAAGCAAG..... 89062
455 euGluThrIleLeuPheGlySerLeuLeuLeuTyrPheProValValIle 471
|||||
89061 .....GTAAACCCAGCAACCTG.....GTTATGATC 89035
472 LeuTyrPheGluProSerThrPheArgCysIleLeuLeuArgTrpAlaAr 488
|||||
89034 CTGTATTAC..... 89026
488 gLeuLeuGlyPheAlaThrValTyrGlyThrValThrLeuLysLeuHisA 505
|||||
89025 .....A 89025
505 rgValLeuLys ValPheLeuSerArgThrAlaGlnArgIleProTyrMe 521
|||||
89024 GAGCAACCTCTTATCTTTTAAAGCAA.....CAGATTTCCTCAATCCA 88981
521 tThrGlyGlyArgValMetArgMetLeuAlaValIle..... 533
|||||
88980 AGTTGCTCCCTCTTAAGCAGCAGCTTTTCTACAATATGATTGGACTAGA 88931
534 ..LeuLeuValValPheTrpPhe..... 540
|||||
88930 ACATGAAGTTATATCTTTAGCTGACCAGAAAATAGAGATTAAACAT 88881
541 .....LeuIleGlyTrpTh 545
|||||
88880 GATACCTTTTAAACATTTACAGACTGTTCCCTGTTGTTAGCCTGGAT 88831
545 rSerSerValCysGlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnG 562
|||||
88830 CTCCTAGTGT.....T 88820
562 lYlYThrSerAspHisLeuIlePheAsnMetCysLeuIleAspArgTrp 578
|||||
88819 CATTGACTTAAATGACTCTCTTTTACATCATCTTTGGCCTCTAGAGAA 88770
579 AspTyrMetThrAlaValAlaGluPheLeuLeuLeuTrpGlyValTy 595
|||||
88769 ACTTATCTACCAGCATCAATGATGATTATTAATTTTGTATCCAAGAA 88720
595 rLeuCysTyrAlaValArgThrValProSerAlaPheHisGluProArgT 612
|||||
88719 GCTTTGTTAT..... 88710
612 yrMetAlaValAlaValHisAsnGluLeuIleIleSerAlaIlePheHis 628
|||||
88709 .....GTTTATTAT 88701
629 ThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTrp...MetLe 644
|||||
88700 APTTTTCTTTCTTAGTGTGTTTAAATA.....TCGTAATGAC 88661
644 uMetLeuTyrPheAlaHisThrHisLeuThrValThrValThrIleGlyL 661
|||||
88660 TCAACTATTTTTCAGGCATATCTTGAAGAAATGAATTTATTTATGACC 88611
661 euLeuLeu...IleProLysPheSerHisSerSerAsnAsnProArgasp 676
|||||
88610 TTATATTGCTTACATAGTGTCCAGAAATATAACAATCCCGGGAAG 88561
```

```

677 AspIleAlaThrGluAlaTyrGluAspGluLeuAspMetGlyArgSerGI 693
88560 AACTTA.....CACAGTTTGGAGAGACATGCTTGTGTTAGTCT 88517
693 ySer_TyrLeuAsnSerIleAsnSerAlaTtp..... 704
88516 AAGTAATCAAGAGTAGTACATCCAGGAGCAGTGCTCATGCCGTGTAATC 88467
705 SerGluHIs.....SerLe 709
88466 TCAGCACATTTGGAGGCCAAGGCTGGAGAATTGCTTGAGGCCAGGAGTTT 88417
709 u_AspProGluAspIleArgAspGluLeuLysLysLeuTyrAlaGlnLeu 725
88416 GAGACACCGCTGACCAACATAGCAAGACCCTGCTCTACAAAAAACA 88367
726 GluIleTyrLysArgLysLysMetIleThrAsnAsn.ProHIsLeuGlnL 742
88366 CAAATTTTAAAGCTAGCAATGCTGCTGATGCTATAGTCCAA 88317
742 ySLysArgCysSerLysLysGlyLeuGlyArgSerIleMetArgIle 758
88316 CTATTTAGGAGCTGAGAAAGAGTA.....TCTCTTGTAGTCTAGAGTT 88273
759 ThrGluIleProGluThrValSerArgGlnCys 769
88272 CAAGACTACAGTACGCTATGATCATCACCACCTGT 88240

seq_name: gb_htg20:AL161654

seq_documentation_block:
LOCUS AL161654 159596 bp DNA HTG 11-APR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-59622, *** SEQUENCING IN
PROGRESS ***, 8 unordered pieces.
ACCESSION AL161654
VERSION AL161654.8 GI:13620309
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Burton,J.
Direct Submission
Submitted (10-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 12, 2001 this sequence version replaced gi:13567947.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA59G22
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 157954 bases at least Q40
Consensus quality: 158428 bases at least Q30
Consensus quality: 158729 bases at least Q20
Insert size: 158896; sum-of-contigs
Quality coverage: 6.69x in Q20 bases; agarose-fp
coverage: 6.67x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

```

* as soon as it is available and the accession number will
* be preserved.
* 1 24276: contig of 24276 bp in length
* 24277 24376: gap of 100 bp
* 24377 31732: contig of 7356 bp in length
* 31733 31832: gap of 100 bp
* 31833 67342: contig of 35510 bp in length
* 67343 67442: gap of 100 bp
* 67443 70588: contig of 3146 bp in length
* 70589 70688: gap of 100 bp
* 70689 81295: contig of 10607 bp in length
* 81296 81395: gap of 100 bp
* 81396 111898: contig of 30503 bp in length
* 111899 111998: gap of 100 bp
* 111999 126181: contig of 14183 bp in length
* 126182 126281: gap of 100 bp
* 126282 159596: contig of 33315 bp in length.
FEATURES
Location/Qualifiers
1..159596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-59G22"
/clone_lib="RPC1-11.1"
1..24276
/note="assembly fragment:01962
fragment_chain:1
clone_end:T7
vector_side:left"
24377..31732
/note="assembly fragment:01615
fragment_chain:1"
31833..67342
/note="assembly fragment:02112
fragment_chain:1"
67443..70588
/note="assembly fragment:00781
fragment_chain:1"
70689..81295
/note="assembly fragment:00892
fragment_chain:1"
81396..111898
/note="assembly fragment:02522
fragment_chain:1"
111999..126181
/note="assembly fragment:01840
fragment_chain:1"
126282..159596
/note="assembly fragment:01914
fragment_chain:1
clone_end:SP6
vector_side:right"
BASE COUNT 54172 a 29956 c 28660 g 46068 t 700 others
ORIGIN

alignment_scores:
Quality: 407.50 Length: 447
Ratio: 1.827 Gaps: 19
Percent Similarity: 49.888 Percent Identity: 30.649

alignment_block:
US-09-775-181-2 x AL161654/rev ..
Align seg 1/1 to reverse of: AL161654 from: 1 to: 159596
372 ArgGlyProAspGlnHIsSerGlySerThrLysAspValSerGluGI 388
102605 AGGGTCCGGATCAGCATATTTAGGAAGTACAAAGATGTGTCAAGA 102556
383 uAlaTyrValCysLeuProCysArgGluGlyCysProPheCysAlaAspA 405
102555 AGCCTATGCTGCTACCTACCTGTCAGGGAGGGTGCCTTCTGTGCTGATG 102506

```

```
405 spSerProCysPheValGlnGluAspLysTyrLeuArgLeuAlaIle 421
|||||
102505 ACAGCCCATCTTGGTGCAGAAAGATTAATACGACTTGGCATCATC 102456
|||||
422 SerPheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeuValVa 438
|||||
102455 TCCTTCCAGCCCTGTGTATGCTGCTCGACTTCGTTAGCATGTGGTGGT 102406
|||||
438 lTyrHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuIleLeuL 455
|||||
102405 CTACCACTTCGCAAGCAAG ..... 102384
455 euGluThrIleLeuPheGlySerLeuLeuLeuTyrPheProValValIle 471
|||||
102383 .....GTAACCCAGCAACCCGCTG .....GTTATGATC 102357
|||||
472 LeuTyrPheGluProSerThrPheArgCysIleLeuLeuArgTrpAlaAr 488
|||||
102356 CTGTATTAC ..... 102348
488 gLeuLeuGlyPheAlaThrValTyrGlyThrValThrLeuLysLeuHisA 505
|||||
102347 .....A 102347
505 rgValLeuLys.ValPheLeuSerArgThrAlaGlnArgIleProTyrMe 521
|||||
102346 GAGCAACCTCTTATCTTTTAAGCAAA .....CAGAAATTTCCAATTCA 102303
|||||
521 tThrGlyGlyArgValMetArgMetLeuAlaValIle ..... 533
|||||
102302 AGTTGCTTCCTCCTAGCAGCAGCTCTTCTACAAATATGATGCGTAGA 102253
|||||
534 ..LeuLeuValValPheTrpPhe ..... 540
|||||
102252 ACATGAAGTTATATTCTTTAGCTGACCAGAAAATAGATTAACAAT 102203
|||||
541 .....LeuIleGlyTrpTh 545
|||||
102202 GATACTTTATTTTAACATTTACAGACTGTTTCCTCCCTGTTTAGCCGCGAT 102153
|||||
545 rSerSerValCysGlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnG 562
|||||
102152 CTCAGTGCT .....T 102142
|||||
562 lLysThrSerAspHisLeuIlePheAsnMetCysLeuIleAspArgTrp 578
|||||
102141 CATTGACTTAAATGACTCTCTTTACATCATGCTTTGGCCTCTAGAGAA 102092
|||||
579 AspTyrMetThrAlaValAlaGluPheLeuPheLeuLeuTrpGlyValTy 595
|||||
102091 ACTTATCTTACCAGCATCCAATGATATTATTAATTTTGTATCCAAGAA 102042
|||||
595 rLeuCysTyrAlaValArgThrValProSerAlaPheHisGluProArgT 612
|||||
102041 GCCTTGTAT ..... 102032
|||||
612 yrMetAlaValAlaValHisAsnGluLeuIleIleSerAlaIlePheHis 628
|||||
102031 .....GTTTATTAT 102023
|||||
629 ThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTrp...MetLe 644
|||||
102022 ATTTTCTTTCTTCTTAGTTGTTTAAATA .....TGCTAAATGAC 101983
|||||
644 uMetLeuTyrPheAlaHisThrHisLeuThrValThrValThrIleGlyL 661
|||||
101982 TCACATATTTTACGCCATATATCTTTGAAGATGAATTTATTATGACC 101933
|||||
661 euLeuLeu...lleProLysPheSerHisSerAsnProArgAsp 676
|||||
101932 TTATATTTTGTCTTACATAGTGTCCAGCAATAATAACAATCCCGGAAG 101883
|||||
```

```
677 AspIleAlaThrGluAlaTyrGluAspGluLeuAspMetGlyArgSerGl 693
|||||
101882 AACTTA ..... 101839
|||||
693 ySer.TyrLeuAsnSerSerIleAsnSerAlaTrp ..... 704
|||||
101838 AGATATCAAGAGTAGTACAAATCCAGAGCAGTGGCTCATGCGCTGAATC 101789
|||||
705 SerGluHis .....SerLe 709
|||||
101788 TCAGCACAATGGGAGCCCAAGGCTGGAGAATTGCTTGAGGCCAGAGATT 101739
|||||
709 u.AspProGluAspIleArgAspGluLeuLysLysLeuTyrAlaGlnLeu 725
|||||
101738 GAGACCAAGCTGACCAACATAGCAAGACCGCTGCTCTACAAAAAACA 101689
|||||
726 GluIleTyrLysArgLysLysMetIleThrAsnAsn.ProHisLeuGlnL 742
|||||
101688 CAATTTTAAAGCTAGCCAAATGGTGGTGCCTATAGTCCCTATAGTCCCA 101639
|||||
742 yLysArgCysSerLysLysGlyLeuGlyArgSerIleMetArgArgIle 758
|||||
101638 CTATTTAGGAGGCTGAGAAAGAGTA .....TCTCTTGAGTCTAGATT 101595
|||||
759 ThrGluIleProGluThrValSerArgGlnCys 769
|||||
101594 CAAGACTACAGTGCATGATCATCACACCACTGT 101562
|||||
seq_name: gb_htg5:AC017271

seq_documentation_block:
LOCUS AC017271 10892 bp DNA HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC017271
VERSION AC017271.1 GI:5553715
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 10892)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10210111 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..10892
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 3383 a 2213 c 2125 g 3171 t
ORIGIN
```

```
alignment_scores:
Quality: 338.50 Length: 566
Ratio: 1.273 Gaps: 25
Percent Similarity: 46.996 Percent Identity: 23.675

alignment_block:
US-09-775-181-2 x AC017271/rev ..
```

Align seg 1/1 to reverse of: AC017271 from: 1 to: 10892

```
226 LeuArgArgLysTrpArg...Pro.HisLeuHisArgArgGlyProAsnG 241
|||||
```



```

9487 CTTTCAGAGTTGGCGGCTGCCTTCATTCGCCCGCAGGAGGATGTGTGC 9438
241 lnglyProArgglyLeuGlyHisSerTrpArgLysAspGlyLeuGly 257
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9437 AA.....CGATGGCTGGAGGAAGTTTCGG..... 9412
258 GlyAspLysSerHisPheLysTrpSerProProtyrLeuGluCysGluAs 274
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9411 .....TCGCGTCATGGGTGATGTGTGCCA 9386
274 nGlySerTyrlsProgly.....TrpLeuValThrLeuSers 287
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9385 AACTTATACGATAATTCGTCACATTAAGTCTCGATAATA..... 9346
287 erAlaIleTyrlsGlyLeuGlnProAsnLeuValProGluPheArgGlyVal 303
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9345 .....TATATAATACCATATCTC..... 9328
304 MetLysValAspIleAsnLeuGlnLysValAspIleAspGlnCysSerSe 320
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9327 .....TCTAGTTG 9320
320 AspGlyTrpPheSerGlyThrHisLysCysHisLeuAsnAsnSerGluC 337
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9319 CGAT.....CGGAATACGACCTCTCGCTGCTCACCAGAACAAAC 9279
337 ysmetProIleLysGlyLeuGlyPheValLeuGlyAlaTyrlsGluCysIle 353
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9278 CCGCGCCACCCGGGATGG.....TACACCTGTCGT 9247
354 CysLysAlaGlyPheTyrlsProGlyValLeuProValAsnAsnPheAr 370
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9246 TGGCGGAATCTTACTACCTGCCCAAT...TCCACGCTCCAGGATTCGG 9200
370 gArgArgGlyProAspGlnHisIleSerGlySerThrLysAspValSerG 387
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9199 T.....GGCGATCGAGTGGAACTGTCCG 9177
387 lu.....GluAlaTyrlsValCysLeuProCysArgGluGlyCysProPhe 401
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9176 AGGCTACGACAACTACGTGTCATTCGTCGCCGCGGATGCACCTAAC 9127
402 CysAlaAspSerProCys.....PheValGlnGlu..... 412
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9126 TGCATAGCAACGCGCTCTCTGACCTTTCAGGAGGAGGTGTCTCAA 9077
413 ....AspLysTyrlsArgLeuAlaIleLeuSerPheGlnGlyLeuCysM 428
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9076 TGTGGCCCTGTCTGCCCTCTCGTGGCCATCGTACTGGCGCCCTGCA 9027
428 etLeuLeuAspPheValSerMetLeuValValTyrlsPheArgLysAla 444
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9026 TCTGTGCTGCATTGCTCTCGGGGTGATTGCTTCGGCAGCGAAGTGC 8977
445 LysSerIleArg..... 448
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8976 AAGGTATTATTTACATATCTTAAGTTGGAACTTTGAGTAATCCTT 8927
449 .....AlaSerGlyLeu...IleLeuLeuGluThrI 458
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8926 TCATCATTTTCAGGCCAATTCGCTGTATGTGGACTGTCTGGAGACGA 8877
458 leLeuPheGlySerLeuLeuLeuTyrlsPheProVal.ValIleLeu..... 472
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8876 TACTGTGGGCAATGTTTACTTATTCATCTGTGACTGTCTCTGGCGAT 8827
472 ..... 472
8826 GGTACCTTCAACTTCTGTGTGATTATACATATTCATCTTTAGGTTGCCG 8777
473 .....TyrlsGluProSerThrPheArgCysIleLeuLeuArgTrpAla 487
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8776 TCCATTTCTTCCCGCATCCACCGAGCGCTTCTGGAGCCCTGGCTT 8727

```

```

488 ArgLeuLeuGlyPheAlaThrValTyrlsGlyThrValThrLeuLysLeuHi 504
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8726 CGGAGCTGGCTTCATCACCTGCTACGGCGCCATCATCTGAAGTTGTA 8677
504 sarGValLeuLysValPheLeuSerArgThrAlaGlnArgIleProTyrm 521
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8676 CCCCACCTGCTGGTACCTTCGCACCCCGGAAAGCGCATCGTGG...GTGC 8630
521 etThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLeuValVal 537
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8629 TGGCGCAGTGGACCTGCTCAAGTATCTGGCACCATGGTCTTTGCCGTC 8580
538 PheTrpPheLeuIleGlyTrpThrSerSerValCysGlnAsnLeuGluLy 554
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8579 ATCTGCTACATGGCGCTTTACGGCTCTGCTGTGGATCTTCTCGAAAG 8530
554 s...GlnIleSerLeuIleGlyGlnGlyLysThrSerAspHisLeuIleP 570
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8529 TGGCAGCTGGAGAGCTCAGGGAGCGGACAG..... 8496
570 heAsnMetCysLeuIleAspArgTrpAspTyrlsMetThrAlaValAlaGlu 586
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8495 ..AACACCTGCCATCCCTCAAGTGGGAAGTGGTCACGACACCGAGCGAG 8448
587 PheLeuPheLeuLeuTrpGlyValTyrlsLeuCysTyrlsAlaValArgThrVa 603
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8447 ATGCTCATCTGCTGTGTGGACTGTCACCTGTCCATCGCCAGTGGGAATGC 8398
603 lProSerAlaPhe..... 607
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8397 CAACACCCAGTTCCGGGTAAAGATGTACAGTCAATCGATTGCAGATCTTT 8348
607 ..... 607
8347 CTTTATCAATTAAAGCGCACAGAGTTGGCTTAGATGATTAACTAAGTAGAA 8298
607 ..... 607
8297 GGGGAATTTAAGTCAAGTCAAAATGAATTAATCGTAATTTCTTTTGTCC 8248
608 ....His.GluProArgTyrlsMetAlaValAlaValHisAsnGluLeuIle 622
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8247 CATGAAGAACGACAAATTTCTGTCACCGCACTGACGCTGGAGTTCTCTG 8198
623 lIleSerAlaIlePheHisThrIleArgPheValLeuAlaSerArgLeuGl 639
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8197 GTCGCTCCAGCTTCTACTTTCGCGCTTTGTCTACCTGCCGGAATGAG 8148
639 nSerAspTrpMetLeuMetLeuTyrlsPheAlaHisThrHisLeuThrValT 656
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8147 TCCAGCGCCATCTTGTGGCCCTGTTTCATCCGCTCCCGCTGACGACGAATA 8098
656 hrValThrIleGlyLeuLeuLeuIleProLys.....PheSerHisSer 670
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8097 GTTTCGCTTGGGTCTGATATTTGTGCCAAAGTTGTGTATCCACGACACAAG 8048
671 SerAsnAsnProArgAspAspIleAlaThrGluAlaTyrlsGlu 684
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8047 CAGTAAC.....GATTGAGTTCCGAACGATATGAT 8015

```

seq_name: gb_in1:AC009845

seq_documentation_block:

LOCUS AC009845 183766 bp DNA INV 24-FEB-2001
DEFINITION Drosophila melanogaster, chromosome 2L, region 25A-25B, BAC clone
BACR04C19, complete sequence.

ACCESSION AC009845
VERSION AC009845

KEYWORDS HTG

SOURCE fruit fly.
Drosophila melanogaster

ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 183766)

Celniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, J., An, H., Baldwin, J., Beeson, K.Y., Busam, D.A.,
Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,
Fierriera, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Hock, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Pacifer, J., Parag, V., Park, S., Patel, S., Pfeiffer, B.,
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Svirskaas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Sequencing of Drosophila chromosome 2L, region 25A-25B

Unpublished

2 (bases 1 to 183766)

Celniker, S.E., Agbayani, A., Arcaluna, T.T., Baxter, E., Blazef, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacifer, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirskaas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.

Direct Submission

Submitted (03-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Feb 24, 2001 this sequence version replaced gi:6532035.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu.

FEATURES

Location/Qualifiers
1. 183766
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="25A-25B"
/clone="BACR04C19 (D1087)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBAC3.6)"

BASE COUNT 52592 a 39362 c 39573 g 52239 t

ORIGIN

alignment_scores:

Quality: 338.50 Length: 566
Ratio: 1.273 Gaps: 25
Percent Similarity: 46.996 Percent Identity: 23.675

alignment_block:

US-09-775-181-2 x AC009845/rev ..

Align seg 1/1 to reverse of: AC009845 from: 1 to: 183766

226 LeuArgArgLysTrpArg...Pro.HisLeuHisArgArgGlyProAsnG 241

11376 CTTTCAGAGTTGCGGGCTGCTTCATTCGCCGACGAGGATGTGTC 11327

241 lnglyProArgGlyLeuGlyHisSerTrpArgArgLysAspGlyLeuGly 257

11326 AA.....CGATGGCTCGAGGAAGTTTTCGG..... 11301
258 GlyAspLysSerHisPheLysTrpSerProTyrLeuGluCysGluAs 274
11300TCGCGCTCATGGTGAGTTGGCCCA 11275
274 nGlySerTyrLysProGly.....TriLeuValThrLeuSerS 287
11274 AACTTTATACGATAATTCGTGACATTAGTTCGGATAATA..... 11235
287 erAlaIleTyrGlyLeuGlnProAsnLeuValProGluPheArgGlyVal 303
11234TATATAATACCATATCTC..... 11217
304 MetLysValAspIleAsnLeuGlnLysValAspIleAspGlnCysSerS 320
11216TGTAGTGTG 11209
320 rAspGlyTrpPheSerGlyThrHisLysCysHisLeuAsnAsnSerGluC 337
11208 CGAT.....CGAATACGACCTTCTGCCTGCACCCGAGAACAAAC 11168
337 ysMetProIleLysGlyLeuGlyPheValLeuGlyAlaTyrGluCysIle 353
11167 CCGCCGCCACCCCGGATGTG.....TACACCTGTCTGTG 11136
354 CysLysAlaGlyPheTyrHisProGlyValLeuProValAsnAsnPheAr 370
11135 TCCCGGGAATCTTACTACCTGCCCAAT...TCCACGCTCCAGGATTCGG 11089
370 gArgGlyProAspGlnHisIleSerGlySerThrLysAspValSerG 387
11088 T.....GGCGATCGAGTGGAACTGTCCG 11066
387 lu.....GluAlaTyrValCysLeuProCysArgGluGlyCysProPhe 401
11065 AGGGTACGACAACTACTCGTTCATTCGTCGCCGCGGATGCACATAAC 11016
402 CysAlaAspAspSerProCys.....PheValGlnGlu..... 412
11015 TCGGATACCAACCGCTGTGTGCTGACCTTTTCAGGAGGAGGAGGTCTCAA 10966
413AspLysTyrLeuArgLeuAlaIleIleSerPheGlnGlyLeuCysM 428
10965 TGTGACGCCCTCTCTCGCCCTCTCGTGGCCCATCTACTTGGSCGCTGCA 10916
428 eLeuLeuAspPheValSerMetLeuValValTyrHisPheArgLysAla 444
10915 TCCTGTGTGCATTCCTCGGGGTGATTGTCTCCGCGACGAAAGTGC 10866
445 LysSerIleArg..... 448
10865 AAGGTATTAGTTTACATATATCCTAAAGTTGGAAACTTTGAGTAATCCTT 10816
449AlaSerGlyLeu...IleLeuLeuGluThrI 458
10815 TCATCATTTTCAGGCCATTGCTGTGTATGTGGACTGTCTGGGAGACGA 10766
458 leLeuPheGlySerLeuLeuLeuTyrPheProVal.ValIleLeu..... 472
10765 TACTCTGGGCATTTGTTTACTTTATGATCTGTGAGTGTCTCTGGCGAT 10716
472 472
10715 GGTCACTTCAACTTCTGTGTATTTACATATATTCATCTTAGTTGCGG 10666
473TyrPheGluProSerThrPheArgCysIleLeuLeuArgTrpAla 487
10665 TCCATTTCTTCCGCATCCACCGAGCGCTGCCTTCTGGAGCCCTGGCTT 10616
488 ArgLeuLeuGlyPheAlaThrValTyrGlyThrValThrLeuLysLeuHi 504
10615 CGGGAGCTGGGCTTCATCACCTGTGTACGGCCCATCATACTACTGAAGTGTGA 10566

```
504 sArgValLeuLysValPheLeuSerArgThrAlaGlnArgIleProTyrM 521
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10565 CGCCACCTGGTGGACATTTCCGCCAGCCGGAAGGCATCGCTGG...GTGC 10519
521 etThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLeuValVal 537
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10518 TGGCGGACGTGGACCTGCTCAAGTATCTGGCCACCATGGCTTTGGCGTC 10469
538 PheTrpPheLeuIleGlyTrpThrSerSerValCysGlnAsnLeuGlu 554
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10468 ATCTGCTACATGGCGCTTTAGCGCCTCGTGGTGGATCTTCGGAAG 10419
554 s...GlnIleSerLeuIleGlyGlnGlyLysThrSerAspHisLeuIleP 570
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10418 TGGCGACCTGGAGACCTCAGGAGGCGGACACG..... 10385
570 heAsnMetCysLeuIleAspArgTrpAspTyrMetThrAlaValAlaGlu 586
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10384 ..AACACCTGCCATCCCTCAAGTGGGAAGTGGTACGACGACGACGAG 10337
587 PheLeuPheLeuLeuTrpGlyValTyrLeuCysTyrAlaValArgThrVa 603
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10336 ATGCTCATCTGCTGCTTGGACTGCACCTGTCCATGCCCGAGTGGAGTGC 10287
603 lProSerAlaPhe..... 607
10286 CAACACCCAGTTCGGGTAAGATGTACAGTCAAAATCGATTGCAGATCTTT 10237
607 ..... 607
10236 CTTTATCAATTAACCGCAAGAGATGGCTTAGATGATTAAGTAGAA 10187
607 ..... 607
10186 GGGGAATTTAAGTCAAGTCAATCAATTAATCGTAAATTTTCCTTTTGTCC 10137
608 ...His.GluProArgTyrMetAlaValAlaValHisAsnGluLeuIle 622
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10136 CATTGAAGGAACGACAAATTTCTGGTACCACGACTGACGCTGGAGTTCCTG 10087
623 lIleSerAlaIlePheHisThrIleArgPheValLeuAlaSerArgLeuG1 639
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10086 GTCTCGTCCAGCTCTACTTCTGCGCTTGTCTACCTGCCGGAATGAG 10037
639 nSerAspTrpMetLeuMetLeuTyrPheAlaHisThrHisLeuThrValT 656
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10036 TCCAGCGCCCATCTGTGCGCCCTGTTTCATCCGCTCCCGACGTGACGAATA 9987
656 hrValThrIleGlyLeuLeuLeuIleProLys.....PheSerHisSer 670
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9986 GTTTCGCTTGGGTCTGATATTTGTGCCAAAGTGTGTGTACGACACCAAG 9937
671 SerAsnAsnProArgAspAspIleAlaThrGluAlaTyrGlu 684
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9936 CAGTATACCC.....GATTTCAGTTCGAAACGATATGAT 9904
```

seq_name: gb_in1:AE003575

```
seq documentation block:
LOCUS AE003575 327209 bp DNA INV 04-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386046 section 16
of 16, complete sequence.
ACCESSION AE003575 AE002638
VERSION AE003575.2 GI:10727324
KEYWORDS HTG
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 327209)
AUTHORS Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
```

Amnatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J., Basu,A., Andrews-Pfannkoch,C., Baldwin,D., Ballew,K.M., Beeson,K.Y., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dunn,P., Dupin,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Eads,R., Eickbush,T.H., Evangelista,C.C., Ferraz,C., Ferreira,S., Fleischmann,W., Foster,C., Gabrielian,A.E., Garq,N.S., Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z., Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,L., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusskern,D.R., Pacle,J.M., Palazzolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svirskaas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A., Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.

The genome sequence of *Drosophila melanogaster* Science 287 (5461), 2185-2195 (2000)

2 (bases 1 to 327209)

Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.

Direct Submission

Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

On Oct 9, 2000 this sequence version replaced gi:7295650.

Location/Qualifiers

1..327209

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/chromosome="2L"

join(<29815..29907,30773..>30955)

/gene="CG11933"

/product="CG35796"

/db_xref="FLYBASE:FBan0011933"

/db_xref="FLYBASE:FBgn0031615"

/evidence=not_experimental

<29815..>30955

/gene="CG11933"

/map="25A3-25A3"

/db_xref="FLYBASE:FBan0011933"

/db_xref="FLYBASE:FBgn0031615"

/evidence=not_experimental

join(29815..29907,30773..30955)

/gene="CG11933"

/note="CG11933 gene product"

/codon_start=1

/db_xref="FLYBASE:FBan0011933"

/db_xref="FLYBASE:FBgn0031615"

/evidence=not_experimental

/protein_id="AAF50997.1"

/db_xref="GI:7295691"

/translation="MAQPFILSPDCHLWTGELGGSGFCGSGCDYLRHHLSPNPQD"

TITLE Direct Submission
JOURNAL Submitted (23-APR-1999) Berkeley Drosophila Genome Project,
University of California Berkeley, Berkeley, CA 94720, USA
COMMENT For further information about this sequence, including its location
and relationship to other sequences, please visit our web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.

FEATURES

Location/Qualifiers
1..3989
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="GH07312"
1..3989
/gene="BcdNA.GH07312"
109..3774
/gene="BcdNA.GH07312"
/codon_start=1
/product="BcdNA.GH07312"
/protein_id="AAD38614.1"
/db_xref="GI:5052568"
/translation="MFRSPFPFASLLFLLLWSTACGRTRAKRSDVYIAGFFPYGDGVE
NSYTCRGVMPKSLALCHNEHGKILANRYLHMWNDQCNAAGVGVKFFDMHSGPN
KVMLEGAACHTVTDIAKASHHLLTQISYADTPMFTQDAFPNFFRVVPSNAPNAP
RLALLKEFNWTRGVYQNEPRYSILPHNHMVADLDAMEVEVETQSFDVNDVAESLKL
REKDVRIILGNENHFAKAFCEAYKLDYGRAYOWLIMATYSTDMNVTQDSCEV
EIALALEGAILVDLLPLSTSGDITVAGTADLEYLVYDRLRGTEYSRFHGTYDGIWA
AALATQYVAEKREDLLTHFDYRVKDWESVLEALRNTSEFGVTGPVYNNERKANIL
INQFOLGOMEKIGEVHSKSHLDLSLGRPVKVGKTPPKDKRLIYIEHSQVNPYIIV
SASASVGIATVFLAFNIKNQRYIKMSPLNNLLIYGCMMYILSILFLGLDIT
LSSVAFFICTARAWILMAGFSFGAMFSTKRWHSIFITDLKLNKVKIKDYQLFMV
VGLAIDAIITWQIAIDFPYRETQLEPLHNIIDDLVLPENECQSEHMTIEFS
IYYAGKLLVFGFALWETRHVSIPALNDSKHIGFSVYNVETCLAGAAISLVLSDR
KDLVFLSFLIFCTTATLCLVFPKLVKELKRNPOGVVDKRVKRVATLRPMKNGRDS
SVCELEQRRLDKVNTNCRKALMEKENELQALIRKLGPEARKWIDGVYCTGGSNVGS
ELEPLNDIIVRLSAPVREMPSTETVETSDVSTSTHVMDSNFSVOSTVMAP
SLPPKKKHRLKEVHSHPADTMPOIOOQLQHQHOOOQHOQOHOQOQOQ
QQQHHHRLKEVHSHPADTMPOIOOQLQHQHOOOQHOQOHOQOQOQ
QTPATRPKYSVSHRNSTNISTQSELSNMGPHSKPSPVAVIKPTASDHRRTSMGA
LKSFNVSQSLDWTHTLSHAKQSPRYASPORCAFHGHGHTYDNPNTSPQRS
VSEKNRKHKPKQKQVCOSETDSEDERDPPNSQPCVQPRKYSRNSNIQHAHHHS
SPNAPDKORSRQKQKQDSIYGASSETELLEGETAILPIFRKLLTEKSPNIRGRSAV
GSPCNISIKDIVEYL"

BASE COUNT 1018 a 1035 c 1035 g 901 t
ORIGIN

alignment_scores:

Quality: 260.50 Length: 1279
Ratio: 0.433 Gaps: 53
Percent similarity: 47.068 Percent identity: 18.452

alignment_block:

US-09-775-181-2 x AF145639

Align seg 1/1 to: AF145639 from: 1 to: 3989

11 CysLeuLeuAlaGlnLeuGlyLeuGlyAlaValGlyAlaSerArgAs 27
||||| ||| ||||| ||| ||||| |||
573 TGCTTTAATGCGCCGAGTGGCCTT...GCTGAAGGAGTTCATTCG 619
27 pProGlnGlyArg.....ProAsp_SerProArgGluArgThrProLys 41
| : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
620 CCAGAGTGGGCACTGTCTACCAAGATGAGCCAGC.....TAT 657
42 GlyLysProHisAlaGlnProGlyArgAlaSerAlaSerArgSerSe 58
: : : : : || :
658 TCCTGCCGCCCAATCACATGTGGTGACCTGGATGGCAGGTCGA 707
58 rAlaProTrpSerArgSer.....ThrAspGlyThrIleLeuAlaGlnL 73
: : : : : || :
708 GTGGTGGAAACCGACAGGTTCGTCAACGATGTGGCTGAATTCGAAGA 757

73 ysLeuAlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGly 89
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
758 AACTGCGGAGAG.....GACGTGAGGATCATCTCGGCAACTTT 798
90 AspSerHisGlnLeuLysArgAlaAsnCysSerGly...ArgTyrGluLe 105
: : : : : || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
799 AACGAGCACTTTCACGCAAGGCAATCTGTGAGGCTTATAAATTTGATAT 848
105 uAlaGlyLeuProGlyLysTrpProAlaLeuAlaSerAlaHisProSerL 122
: : : : : || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
849 GTATGCGACAGCCATCAATGGCTGATCATGGCT..... 882
122 euHisArgAlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnVal 138
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
883ACCTATCCACGAGTGGTGGATGTC 909
139 MetLeuGlnSerAsnLysSerArgGluGlnAsnLeuGlnAspAspLeuAs 155
: : : : : || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
910 ACGCAGCAGCAGGAGTGCAGTGTGGAGGAG..... 939
155 pTrpTyrGlnAlaLeuValTrpSerLeuLeuGluGlyGluProSerIles 172
: : : : : || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
940ATCCTACAGCCCTTGGAGGT..... 960
172 erArgAlaAlaIleThrPheSerThrAspSerLeuSerAlaProAlaPro 188
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
961GCCATTCAGTGGATCTTTGCCCTTGTCCACCACTGGTGAC 1002
189 GlnValPheLeuGlnAlaThrArgGluGluSerArgIleLeuLeuGlnAs 205
: : : : : || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
1003 ATCACAGTGGCTGGCATTACTGCTGATGATCTTGTGGAGTACGACAG 1052
205 pLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThrGluT 222
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
1053 ACTGCCAGGCACT.....GAATATTCCTCC 1075
222 rpPheHisGlyLeuArgArgLys.....TrpArgProHisLeuHisArg 236
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1076 GCTTTTCATGGCTATACCTACGATGGTATCTGGGACGCTGCCCTGGCCATT 1125
237 ArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrpArgArgly 253
: : : : : || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
1126 CAGTATGTGCCGAA.....AAGCGAGA 1148
253 sAspGlyLeuGlyGlyAspLysSerHisPhe.....LysT 265
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
1149 GGATCTGCTA.....ACACATTTGATTATCCGCTGAAGGACT 1186
265 rpSerProTrpTyrLeuGluCys...GluAsnGlySerTyrLysProGly 280
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
1187 GGGAGAGTGTCTCTCTTGGGCTTACGTAATACATCTTCGAGGT... 1233
281 TrpLeuValThrLeuSerSerAlaIleTyrGlyLeuGlnProAsnLeuVa 297
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1234GTGACGGGAGCCGCTGCTTCTACAAACACGAG..... 1266
297 lProGluPheArgGlyValMetLysValAspIleAsnLeuGlnLysVala 314
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
1267CGCAAGGCCCAACATCTCGATCAATCAGTTTC 1297
314 spIleAspGlnCysSerSerAspGlyTrpPheSerGlyThrHisLysCys 330
: : : : : || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
1298 AGCTGGGACAAATGGAAAGATCGGGGAATACAC...TCACAGAAGTCA 1344
331 HisLeuAsnAsnSerGluCysMetProIleLysGlyLeuGlyPheVal 347
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1345 CACTTGGATTTAAGCTTGGGAAACACGATCAATGGTGGG..... 1386
347 uGlyAlaTyrGluCysIleCysLysAlaGlyPheTyrHisProGlyVal 364
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
1386 1386

364 euProValAsnAsnPheArgArgGlyProAspGlnHisLeuSerGly 380
1387AAAACCTCTCCCAAG..... 1401
381 SerThrLysAspValSerGluAlaTyrValCysLeuProCysArgG1 397
1401 1401
397 uGlyCysProPheCysAlaAspAspSerProCysPheValGlnGluAspL 414
1402GATCGCACTTTGATCTACATCGAGCACAGTC 1432
414 ys.....TyrLeuArgLeuAlaIleLeuSerPheGlnGly 425
1433 AGTCATCAACCATATATATGATCGGCTAGTGTTCGTCATTGGA 1482
426 LeuCysMetLeuLeuAspPheValSerMetLeuValValTyrHisPheAr 442
1483 GTGATTATTGCCACAGTTTTTCTGGCCTTAAACATTAAAGTAT.....CG 1526
442 gLysAlaLysSerIleArgAlaSerGlyLeuIleLeuLeuGluThrIleL 459
1527 CAATCAAGATACATCAAGATGTCCAGTCCCATTTGAACAATCTGATCA 1576
459 euPheGlySerLeuLeuLeuTyrPheProValValIleLeuTyrPheGlu 475
1577 TTGGGGCTGTATGATGACCTATTGAGCATCATTTTCTGGGTCTCGAT 1626
476 ProSer.....ThrPheArgCysIleLeuLeuLeuAr 485
1627 ACCACATTAAGTAGTGTGGCAGCTTTTCCCTATATCTGCAGCTCGAGC 1676
485 gTrpAlaArgLeuLeuGlyPheAlaThrValTyrGlyThrValThrLeuL 502
1677 CTGGATCTTGATGGCTGGATTCAGTCTCAGTTTGGAGCCATGTTCTCGA 1726
502 ysLeuHisArgValLeuLysValPheLeuSerArgThrAlaGlnArgIle 518
1727 AGCGTGGCGGGTGATTCATGATATACCGATCTGAAGCTCAATAAG... 1773
519 ProTyrMetThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLe 535
1774 AAGTGATCAGGACTATCAATGTTATGTTGTTGGCGTGTCTTTGGC 1823
535 uValValPheTrpPheLeuIleGlyTrp...ThrSerSerValCysGlnA 551
1824 CATTGATATAGCCATTATACCACTGGCAGATTGCCGATCCCTTTTACC 1873
551 snLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAspHis 567
1874 CGGAACATAACAGTTGGAACTTGTGATCAGCAGAGATATTGATGATGTC 1923
568 LeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrMetThrAlaVa 584
1924 TTGGTGATCCCGAAACGAGTACTGCCAGTCTGAGCACATGACCATATT 1973
584 lAlaGluPheLeuPhe.....LeuLeuTrpGlyValT 595
1974 CGTTAGCATATTATGCTCTACAAGGAGTCTGTGTTGGTTTGGCGCT 2023
595 yrLeuCysTyrAlaValArgThrValPro...SerAlaPheHisGluPro 610
2024 TTTTGGCTGGAACTCGACATGTTTCTATACCGGCTCTGAACGATTC 2073
611 ArgTyrMetAlaValAlaValHisAsnGluLeuIleLeuSerAlaIlePh 627
2074 AAGCATATTGTTTCTCGCTTATAACGTGTTTCATCCTTCTCTGGCGG 2123
627 eHisThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTrpMetL 644
2124 AGCGGTATATCCCTGGTGTCTCGATCGATCGAAGGATTTAGTTTTTGTCT 2173
644 euMetLeuTyrPheAlaHisThrHisLeuThrValThrValThrIleGly 660

2174 TACTCTCGTTTTT.....ATCATTTTTTGTACGACGACACTTTGTGT 2217
661 LeuLeuLeuIleProLysPheSerHisSerSerAsnAsnProArgAspAs 677
2218 TTGGTGTTCGTACCGAAATGTGTGAGCTGAAGCGGAATCCCGAGGGCGT 2267
677 pIleAlaThrGluAlaTyrGluAspGluLeuAspMetGlyArgSerGlys 694
2268 GGTGGCAAAACGCGTTAGGGCCAGCTTGAGACCCATGTCCAAAACGGAC 2317
694 erTyrLeuAsnSerSerIleAsnSerAlaTrpSerGluHisSerLeuAsp 710
2318 GC..... 2319
711 ProGluAspIleArgAspGluLeuLysLysLeuTyrAlaGlnLeuGluI 727
2320CGGGATTCC.....TCGGTGTCCGAATC 2342
727 eTyrLysArgLysLysMetIleThrAsn...AsnProHisLeuGlnLysL 743
2343 GGAGCAACGATTTGGAGATGTAAAGAACACAACTGCCGATTCGGAAGG 2392
743 ysArgCysSerLysLysGlyLeuGlyArgSerIleMetArgIleThr 759
2393 CGCTGATGGAGAGGAGAACGAGCTGCAGGCCCTTAATCCGCAAGCTGGGA 2442
760 GluIleProGluThrValSerArgGln.....CysSe 770
2443CCCGAG.....GCACGAATGATCGATGGGTGACCTGCAC 2480
770 rLysGluAspLysGluGlyAlaAspHisGlyThrAlaLysGlyThrAlaL 787
2481 AGGTGGTCCCAACCTGCTAGCACTGGAGCCCATACTGAACGATGACA 2530
787 euIleArgLysAsnProGluSerSerGlyAsnThrGlyLysSerLys 803
2531 TTGTTAGGCTCTCAGCTCCACCGGTGCTGAGAGATGCCAGCACCACA 2580
804 GluGluThrLeuLysAsnArgValPheSerLeuLysLysSerHisSerTh 820
2581 GAAGTTACCGAGATCAGCTCGCTGGATGTGTGACCTCGACTCATGTGGA 2630
820 rTyrAspHis.....ValArgAspGlnThrGluGluSerSerSerL 834
2631 GATGGATAACTCCTTTGTGCTGCTGAGTCTACAGTGTATGGCCATCGC 2680
834 euProThrGluSerGlnGluGluThrThrGluAsn..... 846
2681 TTCCTCCCAAAAGAAAGCAATCGATTGTAGAGCACCACTCGCATGCC 2730
847SerThrLeuGluSerLeuSerGlyLysLysLeuThrGlnLysLe 861
2731 CCTGCTCCAATATGATCCACCCCTCCAGCAGCACTGCGAGCAGCACTT 2780
861 uLysGluAspSerGluAlaGluSerThrGluSerValProLeuValCysL 878
2781 ACAGCAACATCAGCAGATGCGAGCAGCAGC.....CTGC 2815
878 ysSerAlaSerAlaHisAsnLeuSerSerGluLysLysThrGlyHisPro 894
2816 AGCAGCAGCAACACCGACAGATCAACAGCAACAGCAGCAGCAGCAT 2865
895 ArgThrSerMetLeuGlnLysSerLeuSerValIleAlaSerAlaLysG1 911
2866 CATCATCGCCCATCTGGAGAGAGAACTCGGTGTCCGCTCAGACCGATGA 2915
911 uLysThrLeuGlyLeuAlaGlyLysThrGlnThrAlaGly..... 924
2916 T.....AATATAGCAGCATCACCAGTACGGGGGCAAGCGGAGCG 2956
925ValGluGluArgThrLys..... 930

221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237
:::|||||
901TATCAGGGCAGAGTCTGGCTGGCCAGCGAGCGCTTGGCCAGCT 944
:::|||||
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
|||||
945 CCTCCTGATGCCATGCCACGACTTCCACGCTGGTGGCGGACCATTT 994
|||||
250 rPArgArgLysAspGlyLeuGlyAspLysSerHisPheLysTrpSer 266
|||
995 GCAATTCGCTCTGAAGGC.....TGGGCA 1017
|||
267 ProProTyrLeuGlu.....Cy 272
|||
1018 GATCCCAAGGCTTCGGGGAATTCCTGAAGAAGGTCCATCCCAAGAGTCTG 1067
|||
272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
|||||
1068 TCCACAATGGTTTGGCCAGAGGTTTGGGAAGAACAATTTAACTGCCAC 1117
|||||
285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
|||||
1118 CTCCAAGAAGGTGCAAAAGGACCTTTACCTGTGCACACCTTTCTCGAGAGG 1167
|||
295AsnLeuValProGluPheArgG 302
|||
1168 TCACGAAGAAGTGGCGACAGGTTTACACACAGCTCGACAGCCTTCGGAC 1217
|||
302 LyValMetLysValAspIleAsnLeuGlnLysVal..... 313
|||
1218 CCCCTGTACAGGGGATGAGAACATCAGCAGTGTGAGACCCCTTACATA 1267
|||
313 313
1268 GATTACAGCGCATTTACGGATATCTACAATGTGTACTTAGCAGTCTACTC 1317
|||
314AspIleAspGlnCys..... 318
|||||
1318 CATTCGCCACGCTTGAAGATATATATACCTGCTTACCTGGGAGAGGGC 1367
|||
319SerSerAspGlyTrpPhe 324
|||
1368 TCTTCACCAATGGCTCTGTGTCGACACATCAAGAAGTTGAGGCGTGGCAG 1417
|||
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
|||||
1418 GTCCTGAAGCACCTACGGCACTAAACTTTACAAACAATATGGGGAGCA 1467
|||
336GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
|||||
1468 GGTGACCTTTGATGAGTGTGTGACCTGGTGGGAACATATTCCCATCA 1517
|||
348Gly 348
|||
1518 ACTGGCACCTTCCCGCAGAGATGGCTCCATCGTGTAAAGAAAGTCGGG 1567
|||
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProG1 362
|||||
1568 TATTACACGCTATGCCAAGAGGAGAAAGACTCTTTCATCAACAGGCA 1617
|||
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis.... 377
|||
1618 GAAATCTGTGGAGTGGTGTCCAGGAGGCTTCTCCCAACTGCA 1667
|||
378IleSerGlySerThrLysaspValSerGluGlu..... 388
|||||
1668 GCCGAGACTCGCTGGCAGGGACCAAGGAGGATCATTTAGGGGGAGGCC 1717
|||
389AlaTyrValCysLeuProCysArgGluGly..... 398
|||||
1718 ACCTGCTGCTTTGAGTGTGTGGAGTGTCTCGATGGGAGTATAGTGATGA 1767
|||

399CysProPheCysAlaAspAsp..... 405
|||
1768 GACAGATGCCAGTGCCTGTAAACAAGTCCCGAGATGACTTCTGGTCCAATG 1817
|||
406SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
|||
1818 AGAACACACCTCTCTGGAGGCCATTCCTGTCCTGCTTATCTTTGAGTGGAGT 1867
|||
418LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
|||
1868 GACATAGAACTATCATAGCCATCGCCTTTCTTGGCCTGGGCATCCTCGT 1917
|||
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
|||
1918 GACCTCTTTGTACCCCTCATCTCGTTCGTACCGGACACCCCGTGG 1967
|||
447 LeArgAlaSerGlyLeuLeuLeuGluThrIleLeuPheGlySerLeu 463
|||
1968 TCAATCTCCAGTAGGAGCTCTGCTATATCATCTTGGCTGGTGGTATTTTC 2017
|||
464 LeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheAr 480
|||
2018 CTCGGCTATGTGCTCCCTTTACCCCTCATCGCAACCTACTACCACATC 2067
|||
480 gCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrG 497
|||
2068 CTGCTACTCTCCAGCGCTCTCTAGTTGGCTCTCTCTGCCATGTGCTACT 2117
|||
497 LyThrValThrLeuLysLeuHisArgValLeuLysVal..... 509
|||
2118 CTGCTTTAGTGACCAAAACCAATCGTATGTCACGCACTCTGGCTGGCAGC 2167
|||
510PheLeuSerArgThrAlaG1 516
|||
2168 AAGAAGAAGATCTGCACCGGAGCCAGCATTCATGAGCGCTTGGGCCCA 2217
|||
516 nArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAlaValI 533
|||
2218 AGTGATCATAGCTCCCTCATCTGTAGTTAGTACAGTAACACTAGTGTGA 2267
|||
533 LeuLeuValValPheTrpPheLeuIleGlyTrpThrSerSerValCys 549
|||
2268 CCTGTGATCATC..... 2278
|||
550 GlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAs 566
|||
2279ATGGAGCTCCCATGCCCATTTGTCTACCCGAGATCAAGGA 2322
|||
566 pHisLeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrMetThrA 583
|||
2323 A.....GTCTACCTTATCTGCAATACGACACCTGGGTGTAGTGGCC 2366
|||
583 laValAla...GluPheLeuPheLeuLeuTrpGlyValTyrLeuCysTyr 598
|||
2367 CTGTGGGTTCACATGGACTCTCATCATGAGCTGTACTACTATGCTCTTC 2416
|||
599 AlaValArgThrValProSerAlaPheHisGluProArgTyrMetAlaVa 615
|||
2417 AAGACCGCAACGTGGCGCAACTTCAATGAGCGCTAAATACATCGCCTT 2466
|||
615 laValAlaHisAsnGluIlelleSerAlaIlePheHisThrIleArgp 632
|||
2467 CACCATGTACACTACCTGCATCATCTGGCTGGCTTCTGCCATTTACT 2516
|||
632 heValLeuAlaSerArgLeuGlnSerAspTrpMetLeuMetLeuTyrPhe 648
|||
2517 TT.....GGGAGCAACTACAAGATCATCATCTACCTACCTTC 2548
|||
649 AlaHisThrHisLeuThrValThrIleGlyLeuLeuLeuIlePhe 665
|||
2549 TTCCGGGTGAGCCCTCAGTGTGACCGTGGCCCTGGGTGCATGTTTACTCC 2598
|||
665 oLys.....PheSerHisSerSerAsnAsnProArgAspAspI 678
|||

```

|||||
2599 GAAGATGATCATCATTCATGCCAAACCTGAGAGAACGTCCGCAGTGCCT 2648
678 leAlaThrGluAlaTyrgluAspGluLeuAspMetGly..... 690
2649 TCAGGACC.....TCGTGATGTTGTCGCGATGCAGCTCGGTGATGGC 2689
691 .....ArgSerGlySerTyrgluAsnSerSerIleAsnSerAl 703
2690 AAAGTCGCGTGGCGCTCAACACACTTCTCTCAAC..... 2722
703 aTrpSerGluHisSerLeuAspProGluAspIleArgAspGluLeuLysL 720
2722 ..... 2722
720 ysLeuTyrgluAlaGlnLeuGluIleTyrgluLysArgLysLysMetIleThrAsn 736
2723 .....ATTTCGGAGAAAGAACCCGGGGCAGGG 2752
737 AsnProHisLeuGlnLysLysArgCysSer..... 746
2753 AATGCCAAATTTCTAAGGCAAGTCTGTGTCATGCTGCAACACAGTGAAG 2802
747 LysLysGluLeuGlyArgSerIleMetArgArgIleThrGluIleProG 763
2803 ACAGGCGCCCAAGGAGACAGCAGTGTGGCAGCGCTCTGTGCACGTGA 2852
763 luThrValSerArgGlnCysSerLysGluAspLysGluGlyAlaAspHis 779
2853 AGACCAAGAGAGCGCTGTAAACAA..... 2878
780 GlyThrAlaLysGlyThrAlaLeuIleArgLysAsnProGluSerSe 796
2879 .....ACAGCGTAATCAAA.....CCCCCTCACTAA 2904
796 rGlyAsnThrGlyLysSerLysGluThrLeuLysAsnArgValPheS 813
2905 AAGTTACCAAGGCTGTGGCAAGACCTGTACCTTTTCAGAT..... 2944
813 erLeuLysLysSerHisSerThrTyrgluHisValArgAspGlnThrGlu 829
2944 ..... 2944
830 GluSerSerSerLeuProThrGluSerGlnGluGluThrThrGluAs 846
2944 ..... 2944
846 nSerThrLeuGluSerLeuSerGlyLysLysLeuThrGlnLysLeuLysG 863
2945 .....GCCAGCACCAAGACCCCTTTACAATGTGGAAAG 2975
863 luAspSerGluAlaGluSerThrGluSerValProLeuValCysLysSer 879
2976 AAGAGACAAATACCCCTTCTGTCACTTCACTCCAGCAGCCCTTCT 3025
880 AlaSerAlaHisAsnLeuSerSerGluLysLysThrGlyHisProArgTh 896
3026 ATGGTGTGTGCACCGCGGCGCCACCCGTCGCCACACACACCTCTGCC 3075
896 rSerMetLeuGlnLysSerLeuSerValIleAlaSerAlaLysGluLysT 913
3076 ACCCCATCTG.....ACCGCAGAGAGACCC 3101
913 hrLeuGlyLeuAlaGlyLysThrGlnThrAlaGlyValGluGluArgThr 929
3102 CCTGTCTCTGCTGATTCGCTCATCCCAAGGCTTG..... 3139
930 LysSerGlnLysProLeuProLysAspLysGluThrAsnArgAsnHisSe 946
3140 .....CCTCCTCTCTCCGCGAGCAG..... 3163
946 rAsnSerAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsnP 963
|||||

```

```

3164 .....CCACAGAGCGCCGCC 3179
963 roAlaGluProArgLysProGlnLysSerGlyIleMetLysGlnGln 979
||| ..... |||
3180 CTCAGAGCCCCCGCAGAGCCCAAG.....TCCCTGATGGACAGCTG 3223
980 ArgValAsnProThrThrAlaAsnSerAspLeu..... 990
3224 CAAGGCGTAGTCACCAACTTCGGTTCGGGATTCACGATTTCCATCGGT 3273
991 .....AsnProGlyThrThrGlnMetLysAspAsnPheAspIleGlyC 1005
3274 GCTGGCAGGCCCGGGACACCA.....GGAA 3299
1005 luValCysProTrpGluValTyrgluLeuThrProGlyProValProSer 1021
3300 AC.....AGCCTGGCCTCTCTGTACCGCCCGCTCCGCCG 3337
1022 Glu.....SerLysValGlnLysHisValSerIleValAlaSerGluMe 1036
3338 CAACACCTGCAGATGCTGCCCTGCACCTGAGCACCTTCCAGGAGGATC 3387
1036 tGluLysAsnProThrPheSerLeuLysGluLysSerHisLysProL 1053
3388 CATCTCCCTCTCTGGGAGGACATCGATGACAGT..... 3424
1053 ysAlaAlaGluValCysGlnGlnSerAsnGlnLysArgIleAspLysAla 1069
3425 .....GAGAGATTCAGCTCCTG 3442
1070 GluValCysLeuTrpGluSerGlnGly.....GlnSerIleLeuG 1083
3443 CAGGAGTTCGTGTACGAGCGCGAAGGAACACCGAAGAGATGAATTGGA 3492
1083 uAspGluLysLeuLeuIleSerLysThrProValLeuProGluArgAlaL 1100
3493 AGAGAGGAGGAGCTGCCACAGCAGCAAGCTGACCCCTGAGGATTC 3542
1100 ysGluGluAsnGlyGlyGlnProArgAlaAlaAsnValCysAlaGlyGln 1116
3543 CTGCTCCTGAGCTCTCTCTCTCCGAGATTCGTCGCCCTCTGGCAGC 3592
1117 SerGluGluLeuPro.....LysAlaValAlaSerLysThrGluAs 1131
3593 TCAGTGCCTCAGTTCCTCCCTGATCTGAGTCGGTCTGCACCCCTCCAAA 3642
1131 nGluAsnLeuAsnGlnIleGlyHisGlnGluLysLysThrSerSer 1147
3643 TGTAACCTACGCTCTGTCTGAGGAGCTACAAAGCAAGCTCTCTCC 3691
seq_name: gb_in2:AF318273
seq_documentation_block: 3663 bp mRNA INV 28-FEB-2001
LOCUS AF318273
DEFINITION Drosophila melanogaster metabotropic GABA-B receptor subtype 2
(GABA-B-R2) mRNA, complete cds.
ACCESSION AF318273
VERSION AF318273.1 GI:13160944
KEYWORDS fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 3663)
AUTHORS Mezler,M., Muller,T. and Raming,K.
TITLE Cloning and functional expression of GABA-B receptors from
Drosophila
JOURNAL Eur. J. Neurosci. 13 (3), 477-486 (2001)
PUBMED 11168554
REFERENCE 2 (bases 1 to 3663)
AUTHORS Mezler,M., Mueller,T. and Raming,K.
TITLE Direct Submission

```

JOURNAL Submitted (02-NOV-2000) Bayer Ag., Agricultural Center, Geb. 6240, Alfred Nobel Strasse 50, Monheim 40789, Germany

FEATURES

Source

1. .3663

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

1. .3663

/gene="GABA-B-R2"

1. .3663

/gene="GABA-B-R2"

CDS

/codon_start=1

/evidence=experimental

/product="metabotropic GABA-B receptor subtype 2"

/protein_id="AAK13421.1"

/db_xref="GI:13160945"

/translation="MFRSPFPFASLLFLLLWSTACGRKAKRSDVYIAGFFPYGDGVE NSYTGVRVPSYKLAGHVNHEGKILANYRLHMWNTQCNAAVGKVFDDMMHSGPN KVMLEGAACHTVTDPIAKASKHWHLTQLSYADHPMTKDAFPNEFRVVPNSNAP NAP RLALKEFNWTGVTVYONEPRYSILPHNHMVADLDAMEVEVETOSFYNDVAESLKKL REKDRIILLGNNEHFAKCAKCEAYKLOMYGRAYQNLWATYSTDWNVYQDSECSVE EIALALEGAILVDLLPLSTSGDITVAGITADEYLVEYDRKRGTEYSRFGHYTYDGIWA AALAIQYVAEREDLLTHFDYRVKDWESVFLALRNTSFEVTVGVFYNRERKANIL INQFLOGMEKIGETHSKQSHLDLSLGRPKVWVKTPPKDRTLIYIEHSQVNPITYIV SASAVIGVIATVLAENIKYRNORYIKMSSPHLNLIIVGCMITYLSIIFLGIDTT LSSVAAPVICTARAWILMAGPSLSFGAMFSKTRVHSIFTLKLNKKVIKDYQLFMV VGVLLAIDIAITWQIADPPFYRETKQLEPHENIDVIVPENEYCOSEHMTIFVS IYAYKGLLWFEAGLWETRHVSIPALNDSKHIGFSYVNFETCLAGAAISLVLSRDR KDLVFLLSFFITFTTATLCFLVFKLVEKRNQGVVDRVRATLRPMKSNRDS SVCELEORLDVNTNCRFKALMEKENELQALIRKLGPEARKWIDGVCTCGSNVGS ELEPILNDIVRLSPPTVEMTSVDSVTSTHVEDMNSFVSVOSTVWAPS LPKPKKOSIVEHSHHAPRPMQPIQQQLQHQHQMOQHQHQHQHQHQHQHQHQHQ QQHQHRRHLERNKNSVSAQDDNIGSITSTAGRSKGCDCSMRERKQSTASRHYDSGQ TPTAPKYSVSHRNSSTNISTQSELSNMCHSPKSTPAVIKTPTASDRRTSMGSA LKSNFVSDMLDHTLSHAKQRPSPRYASQPCAEHGHGHTYDNPNTTSPIQRSV SKNNKIRPKPKQKTCQSTEDSERDPPNPQCPQPKVSRSSNTIOHAAHHSS PNVPDKORSRQKODSIYCASSETELLEGETAILPFRKLLTEKSPNYGRSAVG QSCPNTSKDIVEYL"

BASE COUNT 900 a 982 c 968 g 813 t

ORIGIN

alignment_scores:

Quality: 250.00 Length: 1295

Ratio: 0.417 Gaps: 54

Percent Similarity: 46.255 Percent Identity: 18.147

alignment_block:

US-09-775-181-2 x AF318273 ..

align seg 1/1 to: AF318273 from: 1 to: 3663

11 CysLeuLeuAlaGlnLeuGlyLeuGlyAlaValGlyAlaSerArgAs 27

465 TGCCTTTAATGCGCGGACGTGGCTT...GCTCAAGAGTGTCAATTGA 511

27 pProGlnGlyArg.....ProAsp.SerProArgGluArgThrProLys 41

512 CCAGAGTGGCGACGTGTCTACCAAGATGACGCACG.....TAT 549

42 GlyLysProHisAlaGlnGlnProGlyArgAlaSerAlaSerAspSerSe 58

550 TCGCTGCCCCCAATCATCATGTGGCTGACCTGGATGCCATGGAGTCA 599

58 rAlaProTrpSerArgSer.....ThrAspGlyThrIleLeuAlaGlnL 73

600 GGTGTGGAACGACGAGCTTCGTCAACGATGTGGCTGAATCATTTGA 649

73 ysLeuAlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGly 89

650 AACTGCGCGAGAAG.....GACGTGAGGATCATTTGGGCAACTTT 690

90 AspSerHisGlnLeuLysArgAlaAsnCysSerGly...ArgTyrGluLe 105

691 AACGAGCACATTGACCAAGGCAATCTCTGAGGCTTATAAATTGGATAT 740

105 uAlaGlyLeuProGlyLysTrpProAlaLeuAlaSerAlaHisProSerL 122

741 GTATGGCAGAGCCTATCAATGGCTGATCATGGCT..... 774

122 euHisArgAlaLeuAspThrLeuThrHisAlaThrAsnPhLeuAsnVal 138

775ACCTATTCCAGGATGGTGGAAATGC 801

139 MetLeuGlnSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAs 155

802 ACCGAGACAGCAGTCCAGTCTCGAGGAG..... 831

155 pTrpTyrGlnAlaLeuValTrpSerLeuLeuGluGluProSerIleS 172

832ATCGCTACAGCCTTGGAAAGT..... 852

172 erArgAlaAlaIleThrPheSerThrAspSerLeuSerAlaProAlaPro 188

853GCCATTCTAGTGGATCTTTTGCCCTTGTCCACCAGTGGTGAC 894

189 GlnValPheLeuGlnAlaThrArgGluGluSerArgIleLeuLeuGlnAs 205

895 ATCACAGTGGCTGGCATTACTGTGTGATGAGTATCTTGTGGAGTACGACAG 944

205 pLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThrGluT 222

945 ACTCGGAGGCACT.....GAATATTCCC 967

222 rpPheHisGlyLeuArgArgLys.....TrpArgProHisLeuHisArg 236

968 GCITTCATGGCTATACCTACGATGGTATCTGGCAGCTGCCCTGGCCATT 1017

237 ArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrpArgArgLy 253

1018 CAGTATGTGCCGAA.....AAGCGAGA 1040

253 sAspGlyLeuGlyGlyAspLysSerHisPhe.....LysT 265

1041 GGATCTGCTA.....ACACATTTTATTATTCGCGTGAAGGACT 1078

265 rpSerProTrpTyrLeuGluCys...GluAsnGlySerTyrLysProGly 280

1079 GGGAGAGTGTCTTCTTGGAGCTCTAGCTAATACATCTCTCGAGGCT... 1125

281 TrpLeuValThrLeuSerSerAlaIleTyrGlyLeuGlnProAsnLeuVa 297

1126GTGACGGGACCCGCTGCTTCTACAAACACGAG..... 1158

297 lProGluPheArgGlyValMetLysValAspIleAsnLeuGlnLysVala 314

1159CGCAAGGCCAACATCTCTGATCAATCAGTTTC 1189

314 spIleAspGlnCysSerSerAspGlyTrpPheSerGlyThrHisLysCys 330

1190 AGCTGGGACAAATGGAAGATCGGGAATACCAC...TCACAGAGATCA 1236

331 HisLeuAsnAsnSerGlyCysMetProIleLysGlyLeuGlyPheValLe 347

1237 CACTTGGATTTAAGCTTGGGAAACACAGTCAATGGGTGGGG..... 1278

347 uGlyAlaTyrGluCysIleCysLysAlaGlyPheTyrHisProGlyValL 364

1278 1278

364 euProValAsnAsnPheArgArgArgGlyProAspGlnHisIleSerGly 380

1279AAACTCTCTCCCAAG..... 1293

381 SerThrLysAspValSerGluGluAlaTyrValCysLeuProCysArgGl 397

1293 1293

397 uGlyCysProPheCysAlaAspSerProCysPheValGlnGluAspL 414
1294TyrLeuArgLeuAlaIleIleSerPheGlnGly 425
414 ysTyrLeuArgLeuAlaIleIleSerPheGlnGly 425
1325 AGGTCAATCAACCATATATATGTCGCTAGTCTCGGTGTCATGGA 1374
426 LeuCysMetLeuLeuAspPheValSerMetLeuValValTyrHisPheAr 442
1375 GTGATTATGCCACAGTTTCTGGCTTAAACATTAAAGTAT.....CG 1418
442 glysAlaIleSerIleArgAlaSerGlyLeuLeuLeuGluThrIleL 459
1419 CAATCAAGATACATCAAGATGTCAGTCCCATTTGAACAATCTCATCA 1468
459 eupheGlySerLeuLeuLeuTyrPheProValValIleLeuTyrPheGlu 475
1469 TTGTGGCTGTATGATTACCTATTGAGCATCATTTCTGGGTCTCGAT 1518
476 ProSer.....ThrPheArgCysIleLeuLeuAr 485
1519 ACCACATTAAAGTAGTGGCAGCTTTTCCCTATATCTGCACAGCTCGAGC 1568
485 gTrPAlaArgLeuLeuGlyPheAlaThrValTyrGlyThrValThrLeuL 502
1569 CTGGATCTTGATGGCTGGATTTCAGTCTCAGTTTGGAGCCATGTTCTCGA 1618
502 ysLeuHisArgValLeuLysValPheLeuSerArgThrAlaGlnArgIle 518
1619 AGAGTGGCGGTGCATTTCGATTATCCAGCTGAAGCTCAAGCTCAAG... 1665
519 ProTyrMetThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLe 535
1666 AGGTGATCAAGCACTATCAATTGTTTATGTTGGTGGCGTCTTTGGC 1715
535 uValValPheTrpPheLeuIleGlyTrp...ThrSerValCysGlnA 551
1716 CATTGATATAGCCATTATAACCACTGGCAGATTGCCGATCCTTTTACC 1765
551 snLeuGluLysGlnIleSerLeuIleGlyGlnGlyThrSerAspHis 567
1766 GCGAANCTAAACAGTTGGAACCTTGCATCAGAGAATATTGATGATGC 1815
568 LeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrMetThrAlaVa 584
1816 TTGGTGATCCCGAANACAGTACTGCCAGTCTGAGCAGATGCCATATT 1865
584 lAlaGluPheLeuPhe.....LeuLeuTrpGlyValT 595
1866 CGTTAGCATATTATGCTCAAGGAGCTGTTGTTGGTTTGGCGCT 1915
595 yLeuCysTyrAlaValArgThrValPro...SerAlaPheHisGluPro 610
1916 TTTTGGCTGGGAACCTGCAGATGTTCTATACCGGCTCTGAACGATCC 1965
611 ArgTyrMetAlaValAlaValHisAsnGluLeuIleIleSerAlaIlePh 627
1966 AGCATATTGTTTCTCGTTTAAACGTGTTCACTCATCTGCTGGCGG 2015
627 eHisThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTrpMetL 644
2016 AGCGGTATATCCCTGGTCTATCGATCGAAGGATTTAGTTTGTCT 2065
644 euMetLeuTyrPheAlaHisThrHisLeuThrValThrValThrIleGly 660
2066 TACTCTCGTTTTT.....ATCATTTTGTACACAGACCCACTTTGTGT 2109
661 LeuLeuLeuIleProLysPheSerHisSerAsnProArgAspAs 677
2110 TTGGTGTGCTACCCANATTTGGTGAGCTGAAGCGGAATCCCGAGGCGT 2159
677 pLeaIaThrGluAlaTyrGluAspGluLeuAspMetGlyArgSerGlys 694
2160 GTGGCAAAACGGTTAGGGCCACGTTGAGCCCATGTCCAAAAACGGAC 2209
694 eTyrLeuAsnSerSerIleAsnSerAlaTrpSerGluHisSerLeuAsp 710
2210 GC..... 2211
711 ProGluAspIleArgAspGluLeuLysLysLeuTyrAlaGlnLeuGluL 727
2212CGGATTCC.....TCGGTGTGCGAAT 2234
727 eTyrLysArgLysLysMetIleThrAsnAsnProHisLeuGlnLysLysA 744
2235 GGAGCAACGATTGCGAGATCTAAAGAACA.....A 2266
744 rGcysSerLysLysGlyLeuGlyArgSerIleMetArgIleThrGlu 760
2267 ACTGCCGATTCCGA.....AAGGCGCTGATGGAAGGAGAACGAG 2307
761 IleProGluThrValSerArgGlnCysSerLysGluAspLysGluGlyAl 777
2308 CTG....CAGGCTTAATCCCAAGCTGGACCCGAGGACGCAATGGAT 2354
777 aAspHisGlyThrAlaLysGlyThrAlaLeuIleArgLysAsnProProG 794
2355 CGATGGGTGACCTGCACAGT..... 2376
794 luserSerGlyAsnThrGlyLysSerLysGluGluThrLeuLysAsnArg 810
2377GGTCCAACGTGCTAGCACTGGAGCCCATACTGAACGATGAC 2421
811 ValPheSerLeuLys..... 816
2422 ATTGTTAGCTCTCAGCTCCACCGTGGCTCGAGAGATGCCAGCACAC 2471
817SerHisSerThrTyrAspHisValArg. 825
2472 ACTTACCGAGATGACCTCGTGGATAGTGTGACCTCGATCATGTGGAGA 2521
826AspGlnThrGluGluSerSerSerLeu 834
2522 TGGATACTCCTTTGTGCTGGTGCAGTCTACAGTGTGCGCCATCGCTT 2571
835 ProThrGluSerGlnGluGluThrThrGluAsn..... 846
2572 CTCTCCAAAAAGAAAGCAATCGATTGTAGAGCACCACTCCCATGCCCC 2621
847SerThrLeuGluSerLeuSerGlyLysLysLeuThrGlnLysLeuL 862
2622 TGCTCCAACTATGATGCGACCCATCCAGCAGCACTGCAGCAGCACTTAC 2671
862 ysGluAspSerGluAlaGluSerThrGluSerValProLeuValCysLys 878
2672 AGCAACATCAGCAGATGCGAGCAGCAC.....CTGCAG 2706
879 SerAlaSerAlaHisAsnLeuSerSerGluLysLysThrGlyHisProAr 895
2707 CAGCAGCAACACCAGCAGATGCAACAGCAACAGCAGCAGCAGCATCA 2756
895 gThrSerMetLeuGlnLysSerLeuSerValIleAlaSerAlaLysGluL 912
2757 TCATCGCCATCTGGAGAAAGAACTCGGTGTCGCTCAGACCGATGAT. 2805
912 ysThrLeuGlyLeuAlaGlyLysThrGlnThrAlaGly..... 924
2806AATAGGAGCAGCATCAGTACGGCGGCAAGCGGAGCGGA 2847
925ValGluGluArgThrLys..... 930
2848 CGAGACTCTCCAGCATCGGAGAGGCGTCAATCGACCGCTCCAGCA 2897
931SerGlnLysProLeuProLysAspLysGluThrAsnA 943


```

79  ProMetAspValAlaSerTyrLeuTyrThrGlyAsp..... 90
111  |||||:|||||
492  CCAGAACTTGTGCAGCTTTCAATATACCTCAGATTGCTTATCTGCCA 541
91  .SerHisGlnLeuLysArgAlaAsn...CysSerGly.....Ar 102
111  |||||:|||||
542  CAAGCATGAGCTTAAGCGACAAACTCTGTTCAAGTATTTTATGAGAGTT 591
102  gTyrGluLeuAla...GlyLeuProGlyLysTrpProAlaLeuAlaSerA 118
111  |||||:|||||
592  GTACCATCTGATGCACAGCAGCAGCGGCTATGGTGACATGTCACAGCG 641
118  laHisProSerLeuHisArgAlaLeuAspThrLeuThrHisAlaThrAsn 134
111  |||||:|||||
642  CT.....ATAACTGGACCTATGTTT 661
135  PheLeuAsnValMetLeuGlnSerAsnLysSerArgGluGlnAsnLeuGl 151
111  |||||:|||||
662  CTGCTGTACATACTGAAGGAACTACGGTGAA..... 694
151  nAspAspLeuAspTTPtyrGlnAlaLeuValTrpSerLeuLeuGluGly 168
111  |||||:|||||
695  .....GTGGATGGAACCTTCAAGATATGGCAGCCAAGAGGGAT 737
168  lu.....ProSerIleSerArgAla.....Ala 175
111  |||||:|||||
738  CTGCAATTGCACATTCTACAAAGATCTATAGCAATGCTGGTGACGAGCT 787
176  IleThrPheSerThrAspSerLeuSerAlaProAlaProGln..... 189
111  |||||:|||||
788  TTGACAAAGTTACTGGGAAAGCTTCGGAGCCACCTGCCCAAGCAGCAGTG 837
190  .....ValPheLeuGlnAlaThrArgGluGluSer..... 199
111  |||||:|||||
838  GTTGCTGTTTCTGTGAAGGCATGACTGTGAGAGGCTCTTAATGGCTAT 887
199  ..... 199
888  GAGGCGCCTGGGACTTGTGAGAGAAATTTTGTGCTGGCGACTGATGGCT 937
200  .....ArgIleLeuLe 203
111  |||||:|||||
938  GGGCAGACAGGTACGATGTGACAGATGGGTATCAGCGTGAGGCTGTGCGT 987
203  uGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluT 220
111  |||||:|||||
988  GGAATAACAACTCAGCTCCGCTCCTGATGTTAAATGGTTTGTGATGATTA 1037
220  hrGluTrp.....PheHisGlyLeuArg 227
111  |||||:|||||
1038  CTACTTGGAGCTTCGGCCGAAACCAATCACCAGAAATCCATGTTTCAGG 1087
228  ArgLys.....TrpArgProHisLeuHisAr 236
111  |||||:|||||
1088  AATCTTGGCAGCACAGTTTCCAATGCCCTCTGGAAGGTTTCCCTCAAGAG 1137
236  g.....ArgG 238
111  |||||:|||||
1138  ATCTTAATACAACAGACTTGCACACGCCAGATGACACTGAGGACCGCA 1187
238  lyProAsnGlnGlyProArg.GlyLeu..... 246
111  |||||:|||||
1188  GCACGTTCCAGGACTCCAAGATGGGCTTTGTAATCAACGCTATATACTCAA 1237
246  ..... 246
1238  TGGCTTACGGCTCCACAACTATGACAGCTGTCTACTGTGCCAGGCTATGTG 1287
247  .....GlyHisSerTrp.ArgArgLys..... 253
111  |||||:|||||
1288  GGCCCTTTCGATGCCATGAGCCCATAGATGGACGGAAGCTTCTGGAGTC 1337
254  .....AspGlyLeuGlyAspLysSerHisPhe. 263
111  |||||:|||||
1338  CCTCATGAAGACCAACTTTACTGGGTCTCTGGGACATGATCTTCTTTG 1387
264  .....Lys 264
111  |||||:|||||
1388  ATGAGAAATGGTGACTGCGCAGGAAGATATGAATCATGAATTTTAAGAAA 1437
265  TrpSerProTyrLeuGluCysGluAsn...GlySerTyrLysProGl 280
111  |||||:|||||
1438  ATCGGAAGGACTACTTTCATTTATATCAATGTTCGACGCTGGCACAACGG 1487
280  yTrpLeuValThrLeuSerSerAlaIleTyrGlyLeuGlnProAsnLeuV 297
111  |||||:|||||
1488  TGAGCTGAAAATGGATGATGATAAATATGGTCAGAGAAAAATAATATCA 1537
297  al.....ProGluPheArgGlyValMetLysValAsp 307
111  |||||:|||||
1538  TCAGATCTGTGTGCACTGAACCTGTGNAAGAGGACAGATAAGGTG... 1584
308  IleAsnLeuGlnLysValAspIleAspGlnCysSerSerAspGlyTrpPh 324
111  |||||:|||||
1585  .....ATTCGTAAGGAGAGTGTCTGT.....TGG.. 1614
324  eSerGlyThrHisLysCysHisLeuAsnAsnSerGluCysMetProIleL 341
111  |||||:|||||
1615  .....ACATCCACTCCCTGTGA 1630
341  ysGlyLeuGlyPheValLeuGlyAlaTyrGluCys...IleCysLysAla 356
111  |||||:|||||
1631  AAGAGAATGAGTACGTCTTTGATGATACACGTGCAAGGCTGCCAGCTC 1680
357  GlyPheTyrHisProGlyValLeuProValAsnAsnPheArgArgGl 373
111  |||||:|||||
1681  GGCTCTCTGG..... 1689
373  yProAspGlnHisIleSerGlySerThrLysAspValSerGluGluAla 390
111  |||||:|||||
1690  .CCCAATGATGAGCTTACAGT..... 1710
390  yValCysLeuProCysArgGluGlyCysPropheCysAlaAspSer 406
1710  ..... 1710
407  ProCysPheValGlnGluAspLysTyrLeuArg..... 417
1711  ...TGTGACCTCATCCAGTCCAGTATCTCAGATGGGTGATCCTGAACC 1757
418  .LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLeuAspPheVal 434
111  |||||:|||||
1758  AATTCACAGCAGTTGTTTTCGCTGCTGGTCTGCTGGCAACCTTGTGTTG 1807
434  erMetLeuValValTyrHisPheArgLysAlaLysSerIleArgAlaSer 450
111  |||||:|||||
1808  TTACGGCTATATTCATAATGTACCGCAGTACTCCAGTGGTCAATCTTCC 1857
451  GlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeuLeuTyrPh 467
111  |||||:|||||
1858  ACCCAGAACCTTGTACATCTTCTAGTGGCATCTGCTGGGTACTT 1907
467  eProValValIleLeuTyrPheGluProSerThrPheArgCysIleLeu 484
111  |||||:|||||
1908  GTGCACTTCTGCTCATCGCAAAACCTCAACAGATTTACTGCTATCTTC 1957
484  euArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrGlyThrValThr 500
111  |||||:|||||
1958  AACGAATTGGCATCGGTCTCTCCCGAGCTATGAGTTATCTGCTCTAGTA 2007
501  LeuLysLeuHisArgValLeuLysVal..... 509
111  |||||:|||||
2008  ACTAAACCAACCGCATTCGACAGAAATCTGGCTGGCAGCAAGAAAT 2057
510  .....PheLeuSerArgThrAlaGlnArgIleProt 520

```

2058 TTGTCACAAAGAACCTAGATTCATGAGTGCCTGTGCCCAA..... 2097
520 yrMetThrGlyArgValMetArgMetLeuAlaValIleLeuLeuVal 536
|||||
2098CTTGT 2103
537 ValPheTrpPheLeuIleGlyTrpThrSerSerValCysGlnAsnLeuG1 553
||| :|||:|||||
2104 AATGCAATTTATCTTGATA.....TCTATACAAATTAGG 2135
553 uLysGlnIleSerLeu.....IleGlyGlnGlyL 563
|||||:|||||
2136 CATATATGTGGCCCTCTTTATAATAGGACCCACCTGATATAATGCATGATT 2185
563 ysThr-SerAspHisLeuIlePheAsnMetCysLeuIleAspArgTrpAsp 579
||| :|||:|||||
2186 ACCCAGCATCGAGAGGTCTACTTGATTGTACACCAACCACTGGGT 2235
580 TyrMetThrAlaValAla...GluPheLeuPheLeuLeuTrpGlyValty 595
||| :|||:|||||
2236 GTTGTAAACCCCTCTGGATACAAATGATTAATTTAGTTGACACCTT 2285
595 rLeuCysTyrAlaValArgThrValProSerAlaPheHisGluProArgT 612
||| :|||:|||||
2286 TTATGCATTTAAGACAAGAAATGTCCAGCTAAATTTCAATGAAGCAAAAGT 2335
612 yrMetAlaValAlaValHisAsnGluLeuIleIleSerAlaIlePheHis 628
||| :|||:|||||
2336 ATATGTCCTTACAAATGATACCACTGCATCATTTGGCTGGCTTTGTG 2385
629 ThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTrpMetLeuMe 645
||| ||| :|||:|||||
2386 CCAATATATTTT.....GGAAGCAACTACAAGATAAT 2417
645 tleuTyrPheAlaHisThrHisLeuThrValThrValThrIleGlyLeuL 662
: |||:|||||:|||||:|||||
2418 CACCATGTGTTTCTCAGTCAGTCTCAGTGCCACGGTGGCCCTTGTTGTA 2467
662 euLeuIleProLys.....PheSerHisSerSerAsnAsnPro 674
:|||||:|||||
2468 TGTGTTGCCCAGGCTCATCATCTCTGTGTAAGCCTGAAGGNACGTA 2517
675 ArgAspAspIleAlaThrGlu...AlaTyrGluAspGluLeuAspMetG1 690
|||:|||||:|||||:|||||
2518 CGCAGCGCATTCACCACTTCGACTCTGTCGCGCATGCTAGTGGGATGG 2567
690 yArgSerGlySerTyrLeu.....AsnSerSerIleAsnSerAlaTrps 705
|||:|||||:|||||:|||||
2568 AAAATCATCTTCGGCTGGAGCGCTCAAGCAGCCCTGTGTGAACCTGTGGA 2617
705 erGluHisSerLeuAspProGluAspIleArgAspGluLeuLysLysLeu 721
|||:|||||
2618 AAAGAAGGGATCATCTGGTGAACACTAGG..... 2649
722 TyrAlaGlnLeuGluIleTyrLysArgLysLysLysMetIleThrAsnAsnPr 738
||||| :|||:|||||
2650TACAAGGCGAGGAGACTGGCC.....CC 2672
738 ohisLeuGlnLysLysArgCys...SerLysLysGlyLeu..... 750
||| :|||:|||||
2673 GCACAAGTCGGAATAAGATGTTTCCACCCCAAAAGGAGTATGGGGAATG 2722
751 ..GlyArgSerIleMet.....ArgArgIleThrGluIleProGluThr 764
|||||:|||||
2723 GTGGAGAGCTACATACCACTGAAGAAATCTGTTTGGCATCTCCAGATGT 2772
765 valSerArgGlnCysSerLysGluAspLysGluGlyAlaAspHisGlyTh 781
: |||:|||||:|||||
2773 AATCGATCTGAGTCAAGTAGAGATGAGAAAGAG..... 2805
781 rAlaLysGlyThrAlaLeuIleArgLysAsnProProGluSerSerGlyA 798
||| :|||:|||||

CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 368 BP; 115 A; 78 C; 79 G; 96 T; 0 other;

alignment_scores:
Quality: 580.00 Length: 112
Ratio: 5.179 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-775-181-2 x AAA42817 ..
Align seg 1/1 to: AAA42817 from: 1 to: 368

626 IlePheHisThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTr 642
|||||
24 ATATTCACATACATAGATTGTTCTTGCTCAAGACTTCAGTCTGATG 73

642 pMetLeuMetLeuTyrPheAlaHisThrHisLeuThrValThrI 659
|||||
74 GATGTTGATGCTGTATTTGCACATACTCATTTGACTGTGACAGTCACCA 123

659 leGlyLeuLeuLeuIleProLysPheSerHisSerAsnAsnProArq 675
|||||
124 TTGGTGTGCTTTGATTCACAAAGTTTTCACATTCAGCAATTAACCCACCA 173

676 AspAspIleAlaThrGluAlaTyrGluAspGluLeuAspMetGlyArgSe 692
|||||
174 GATGATATTGCTACAGACATATGAGGATGAGCTAGACATGGCGCGATC 223

692 rGlySerTyrLeuAsnSerSerIleAsnSerAlaTrpSerGluHisSerL 709
|||||
224 TGATCCTACCTGAACAGCAGATATCAATTCAGCCTGGAGTGACACAGCT 273

709 euAspProGluAspIleArqAspGluLeuLysLysLeuTyrAlaGlnLeu 725
|||||
274 TGGATCCAGGACATCCGGGACGAGCTCAAAAACCTCTATGCCCACTG 323

726 GluIleTyrLysArgLysLysMetIleThrAsnAsn 737
|||||
324 GAATATATATAAAGAAAGAGATGATCATCAAAACAAC 359

seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAA41583

seq_documentation_block:

ID AAA41583 standard; cDNA: 182 BP.

AC AAA41583;

DT 21-AUG-2000 (first entry)

DE Human secreted expressed sequence tag SEQ ID NO:323.

XX Human: mouse; xenopus; rat; secreted expressed sequence tag; sEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; cytostatic; antiparkinsonian;
KW antitumor; osteopathic; neuroprotective; neurotropic; antipsoriatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune diabetes; asthma; myeloid cell deficiency; ulcer;
KW insulin dependent diabetes; multiple sclerosis; allergic condition;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.

XX

OS Homo sapiens.
XX WO200021990-A1.
XX PD 20-APR-2000.
XX PF 15-OCT-1999; 99WO-US24205.
XX PR 15-OCT-1998; 98US-0104435.
XX (GEMY) GENETICS INST INC.
XX PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M;
XX WPI: 2000-317937/27.
XX Isolated polynucleotides, and encoded proteins, comprising secreted
XX expressed sequence tags (seSTs), useful for treating various disorders -
XX such as autoimmune, infectious, and central nervous system disorders -
XX Claim 1; Page 243; 618pp; English.

XX AAA41261 to AAA43419 represent specifically claimed secreted expressed
XX sequence tags (seSTs), isolated from human, mouse, xenopus and rat
XX tissue sources. The seSTs can have a range of activities depending on
XX the tissues they were isolated from. The activities include:
XX chemotactic; proliferative; immunomodulatory; haematopoietic;
XX chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
XX cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
XX antiasthmatic; vulnery; antitumor; osteopathic; neuroprotective;
XX neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective;
XX anticonvulsant; and antidepressant. The seSTs can be used for gene
XX therapy and in vaccines. The seSTs are useful as probes for the
XX identification and isolation of full-length cDNAs and genomic DNA
XX molecules which correspond to the seSTs. Proteins encoded by the seSTs
XX are useful in assays for determining biological activity and raising
XX antibodies. They may be useful for treatment of autoimmune disorders
XX (multiple sclerosis, insulin dependent diabetes), allergic conditions
XX (asthma), myeloid or lymphoid cell deficiencies, wounds, ulcers,
XX osteoporosis, osteoarthritis, central nervous system disorders
XX (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
XX disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
XX disease), tumours, bacterial, fungal or viral infections, depression and
XX psoriasis. AAA43420 to AAA43425 represent linker variants which are given
XX in the exemplification of the present invention.
XX Sequence 182 BP; 46 A; 36 C; 49 G; 51 T; 0 other;

alignment_scores:

Quality: 260.00 Length: 51

Ratio: 5.200 Gaps: 0

Percent Similarity: 98.039 Percent Identity: 98.039

alignment_block:

US-09-775-181-2 x AAA41583

Align seg 1/1 to: AAA41583 from: 1 to: 182

518 IleProTyrMetThrGlyArgValMetArgMetLeuAlaValIleLe 534

|||||

27 ATTCATATATGACTGGCGGAGGGTTCATGAGGATGCTGGCAGTAATACT 76

534 uLeuValValPheTrpPheIleGlyTrpThrSerValCysGlnA 551

|||||

77 CTGTGAGTGTGTTTGGTCTTCATTTGGCTGGACTTCATCTGTGTGCCAGA 126

551 snLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAspHis 567

|||||

127 ATTTGGAGAAACAGATTTCACATTATTGGCCAGGGGAAAACCCGATCAC 176

568 Leu 568

177 CTC 179

seq_name: /SIDS1/gcdata/geneseq/geneseqn/NA1997.DAT:AAT86166

seq_documentation_block:

ID AAT86166 standard; DNA: 4000 BP.

XX AC AAT86166;

XX DT 02-DEC-1997 (first entry)

XX DE Nucleotide sequence of pCar/R1.

XX KW Chimeric receptor; extracellular domain; seven transmembrane domain;
intracellular cytoplasmic tail domain; metabotropic glutamate receptor;
MGR; calcium receptor; CR; mclur; G protein-coupled receptor;
neurological disease; ss.

XX OS Chimeric - Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 41..3700

FT FT /*tag= a

FT FT /product= pCar/R1

XX XN MO9705252-A2.

XX XX 13-FEB-1997.

XX XX 25-JUL-1996; 96WO-US12336.

XX XX 26-JUL-1995; 95US-0001526.

XX XX (NPSP-) NPS PHARM INC.

XX XX Fuller FH, Hammerland LG, Krapcho KJ;

XX XX WPI; 1997-145690/13.

XX XX P-PSDB; AAW25763.

XX PT Chimeric receptors comprising metabotropic glutamate receptor and
calcium receptor - used for screening for neurologically active
compounds

XX PS Example 3; Fig 3; 177pp; English.

XX CC This sequence encodes a chimeric receptor. The chimeric receptor
comprises an extracellular domain, a seven transmembrane domain, and
an intracellular cytoplasmic tail domain, and a sequence of at least
6 contiguous amino acids is homologous to a sequence of a metabotropic
glutamate receptor (MGR), and a sequence of at least 6 contiguous amino
acids is homologous to a sequence of a calcium receptor (CR). The
chimeric receptor may be used for screening for compounds that bind to
or modulate the activity of MGR or CR and for determining the site-of-
action of a CR active compound. The compounds can be used in the
treatment of neurological diseases and disorders. They can also be
used as diagnostic agents. Chimeric receptors such as this, allow the
coupling of certain functional aspects of an MGR with certain functional
aspects of a CR. They allow for more efficient high-throughput screening
of compounds.

XX SQ Sequence 4000 BP; 955 A; 1128 C; 1010 G; 907 T; 0 other;

alignment_scores:

Quality: 254.00 Length: 1285
Ratio: 0.490 Gaps: 48
Percent Similarity: 40.311 Percent Identity: 18.599

alignment_block:

US-09-775-181-2 x AAT86166 ..

Align seg 1/1 to: AAT86166 from: 1 to: 4000

```

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
434 TCAGAGCACATTCCTCTACGATTGCTGTGGGAGCACTGGCTCAGG 483
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
-91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
484 CGTCTCCACGGCAGTGGCAATCTCTGGGGCTCTCTACATT..... 526
108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
527 .....CCCAGGTCAGTTATGCTCTCTCCAGC.....AGA 556
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG1 141
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
557 CTCCTCAGCAACAAGATCAATTCAAGTCTTCTCCGAACC.....ATCCC 603
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAspTrpTyrG 158
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
604 CAATGATGAGCAGCAGGCACTGCGCATGGCAGACATCATCGAGTATTTC 653
158 InAlaLeuValTrpSerLeuLeuGluGlyGluProSerIleSer..... 172
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
654 GCTGGAACTGGTGGGCACAAATTCGAGCTGATGACGACTATGGGGCGCG 703
173 .....ArgAlaAlaIleThrPheSerT 180
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
704 GGGATTGAGAAATTCAGAGAGGAGCTGAGGAAGGGATATCTGCATCGA 753
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
754 CTTCACTGAACATCTCTCCAGTACTCTGATGAGGAGAGATCCAGCATG 803
196 .....ArgGluGluSer.ArgIleLeuLeuG 204
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
804 TGGTAGAGGTGATTCAAAATTCACGGCCAAAGTCATCGTGGTTTCTCC 853
204 InAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
854 AGTGGCCAGATCTTGAGCCCTCATCAAGGAGATTGTCCGGCGCAA... 900
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
901.....TATCAGGGCAAGATCTGGCTGGCCAGCGAGGCTGGCCAGCT 944
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
945 CCTCCCTGATCGCCATGCCCTCAGTACTTCCACGTGGTGGCGGACCAT 994
250 rPArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSer 266
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
995 GGATTCGCTCTGAAGGC.....TGGGCA 1017
267 ProProTyrLeuGlu.....Cy 272
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1018 GATCCAGGCTTCGGGAATTCCTGAAGAGGTCCATCCAGGAAGTCTG 1067
272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1068 TCCACAATGGTTTTCGCAAGAGGATTTTGGGAAGAACAATTTAACTGCCAC 1117
285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1118 CTCCAAGAAGGTGCAAAAGGACCTTTACCTGGACACACCTTTCTGAGAGG 1167
295 .....AsnLeuValProGluPheArgG 302
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1169 TCACCAAGAAAGTGGCAGACAGGTTTAGCAACAGCTCGACACGCTTCCGAC 1217
302 LyValMetLysValAspIleAsnLeuGlnLysVal..... 313
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1218 CCCTCTGTACAGGGATGAGAACATCAGCAGTGTGAGACCCCTTACATA 1267

```

313 313
1268 GATTACAGCATTTACGGATATCTACAATGTGTACTTAGCAGTCTACTC 1317
314AspileaspGlnCys..... 318
1318 CATTTGCCACGCCCTTGCAGATATATATACCTGCTTACCTGGGAGAGGC 1367
319SerSerAspGlyTrpPhe 324
1368 TCTTCACCAATGGCTCTCTGTCAGACATCAAGAAAGTTGAGGGCTGGCAG 1417
325 SerGlyThrHisLysCysHisLysLeuAsnSer..... 335
1418 GTCTGAAGCACCTAGCGCATCTAAACTTTACAAACAATATGGGGAGCA 1467
336GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
1468 GGTGACCTTTGATGAGTGTGGTACCTGGTGGGAACTATTTCATCATCA 1517
348Gly 348
1518 ACTGGCACCTCTCCCGAGGATGGCTCCATCGTGTAAAGGAAGTCGGG 1567
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProGln 362
1568 TATTACAAGCTCTATGCCAAGAGGGAGAAAGACTCTTCATCAACGAGGA 1617
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis..... 377
1618 GAAATCCCTGTGGAGTGGGTTCACAGGAGGTGCCCTTCTCCAACGTCA 1667
378IleSerGlySerThrLysAspValSerGluGlu..... 388
1668 GCCGAGACTGCCCTGGCAGGACCAGGAAGGATATTGAGGGGAGSCC 1717
389AlaTyrValCysLeuProCysArgGluGly..... 398
1718 ACCTGCTGTTTGTGAGTGTGGAGTCTCTGTGATGGGAGTATAGTGATGA 1767
399CysProPheCysAlaAsp..... 405
1768 GACAGATGCCAGTCCCTGTAAAGTGGCCAGATGACTTCTGTGTCCAATG 1817
406SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
1818 AGAACACACCTCTCGGAGCCCATTCCTGTCGTTATCTTGAGTGGAGT 1867
418LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
1868 GACATAGAATCTATCATAGCATCGCTTTCTTGCTGGCATCTCGT 1917
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
1918 GACCTGTTTGTACCCCTCATCTTCCTGTACCCGGGACACACCCGTGG 1967
447 leArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeu 463
1968 TCAATCCTCCAGTAGGAGCTCTGTATATCATCTTCTGCTGGTATTTC 2017
464 LeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheAr 480
2018 CTGGGTATGTGTGCTTTCACCTCTCATCGCCCAACCTACTACCACATC 2067
480 gCysIleLeuLeuLeuTyrAlaArgLeuLeuGlyPheAlaThrValTyrG 497
2068 CTGCTACCTCCAGCGCTCTAGTTGGCTCTCTCTGCTGCTGCTACT 2117
497 lyThrValThrLeuLysLeuHisArgValLeuLysVal..... 509
2118 CTGCTTTAGTACCAAAACCAATCGTATTGACCGCATCTCGCTGGCAGC 2167
510PheLeuSerArgThrAlaGln 516
2168 AAGAGAAGATCTGCACCGGAAGCCAGATTTCATGAGCGCTTGGGCCCA 2217
516 nArgIleProTyrMetThrGlyArgValMetArgMetLeuAlaValI 533
2218 AGTGATCATAGCTCCATTCCTGATTTAGTGTACAGCTAACACTAGTGGTGA 2267
533 leLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerSerValCys 549
2268 CCTGTGATCATC..... 2278
550 GlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAs 566
2279ATGGAGCTCCCATGCCCATTTTGTCTACCCGAGTATCAAGGA 2322
566 pHISLeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrMetThra 583
2323 A.....GTCTACCTTATCTGCAANTACCAACACCTGGGTGTAGTGGCCC 2366
583 laValAla...GluPheLeuLeuTrpGlyValTyrLeuCysTyr 598
2367 CTGTGGTTACAATGGACTCTCTCATCATGAGCTGTACTACTATGCTTC 2416
599 AlaValArgThrValProSerAlaPheHisGluProArgTyrMetAlaVa 615
2417 AAGACCCGCAACGTGCGGCCAACTTCAATGAGGCTAAATACATCGCCT 2466
615 lAlaValHisAsnGluLeuIleIleSerAlaIlePheHisThrIleArgP 632
2467 CACCATGTACACTACCTGTCATCTGGCTGGCTTTCTGCCATTACT 2516
632 heValLeuAlaSerArgLeuGlnSerAspTrpMetLeuMetLeuTyrPhe 648
2517 TT.....GGGAGCACTACAAGATCATCTACTACTGC 2548
649 AlaHisThrHisLeuThrValThrIleGlyLeuLeuLeuLeuLeu 665
2549 TTCGCGGTGAGCTCAGTGTACGGTGGCTGGGTGTCATGTTTACTCG 2598
665 oLys.....PheSerHisSerSerAsnAsnProArgAspAspI 678
2599 GAAGATGTACATCATCTTGCACAACTGAGAGAACTCCCGAGTGCCT 2648
678 leAlaThrGluAlaTyrGluAspGluLeuAspMetGly..... 690
2649 TCAGGACC.....TCTGATGTTGTCGCGCATGCACGTCGCTGATGC 2689
691ArgSerGlySerTyrLeuAsnSerSerIleAsnSerAl 703
2690 AAAGTCCGCTGCCGTCCCAACACCTCTCTCAAC..... 2722
703 aTrpSerGluHisSerLeuAspProGluAspIleArgAspGluLeuLysL 720
2722 2722
720 ysLeuTyrAlaGlnLeuGluIleTyrLysArgLysLysMetIleThrAsn 736
2723ATTTCGGAGAAAGAACCCCGGCGCAGG 2752
737 AsnProHisLeuGlnLysLysArgCysSer..... 746
2753 AATGCCAATTCTAACGGCAAGTCTGTGTCATCTGTAACACAGGTGGAAG 2802
747 LysLysGlyLeuGlyArgSerIleMetArgArgIleThrGluIleProG 763
2803 ACAGGGCCCAAGGACACGACGTGTGGCAGCGCTCTCTGTCAGGTGA 2852
763 luThrValSerArgGlnCysSerLysGluAspLysGluGlyAlaAspHis 779
2853 AGACCAACGAGACGGCTGTACCAAC..... 2878
780 GlyThrAlaLysGlyThrAlaLeuIleArgLysAsnProGluSerSe 796


```
510 .....PheLeuSerArgThrAlaG 516
2693 CAAGAAGAAGATCGCACCGGAGCCAGCATTCATGAGCGCTTGGGCC 2742
516 lnArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAlaVal 532
2743 AAGTGATCATAGCTCCATTCTGTTAGTGTACAGCTAACACTAGTGGTG 2792
533 IleLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerValCy 549
2793 ACCTTGATCATC..... 2804
549 sGlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerA 566
2805 .....ATGAGCGTCCCATGCGCATTTTGTCTACCCGAGTATCAAGG 2847
566 sPHisLeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrMetThr 582
2848 AA.....GTCTACCTTATCTGCAATACCAACACCTGGGTGTAGTGGCC 2891
583 AlaValAla...GluPheLeuPheLeuLeuTrpGlyValTyrLeuCystY 598
2892 CCTGTGGTTACAATGACCTCTCATCATGAGCTGTACCTACTATGCGTT 2941
598 rAlaValArgThrValProSerAlaPheHisGluProArgTyrMetAlav 615
2942 CAAGACCGCAACGTCGCGCAACTTCAATGAGGCTAAATATACATCGCCT 2991
615 alAlaValHisAsnGluLeuIleLeuSerAlaIlePheHisThrIleArg 631
2992 TCACCATGTACACTACCTGCATCATCTGCGTGGCTTTCGTTCCTCCATTAC 3041
632 PheValLeuAlaSerArgLeuGlnSerAspTrpMetLeuMetLeuIleP 648
3042 TTT.....GGGAGCAACTACAAGATCATCACTACTG 3073
648 eAlaHisThrHisLeuThrValThrValThrIleGlyLeuLeuLeuIleP 665
3074 CTTGCGGTGAGCGTCACTGTGACGGTGGCCCTGGGGTGCATGTTTACTC 3123
665 rOlyLys.....PheSerHisSerAsnAsnProArgAspAsp 677
3124 CGAGATGTACATCATCTATGCGCAACCTCAGAGGAACGTCGCGAGTGCC 3173
678 IleAlaThrGluAlaTyrGluAspGluLeuAspMetGly..... 690
3174 TTCAGGACCTCTGATGTGTC...CGCATGAGCGTCGGTGTATGGCAACT 3220
691 .....ArgSerGlySerTyrLeuAsnSerSerIleAsnSerAlaTrpS 705
3221 GCGGTGCGCGCTCCAACACCTTCCTCAAC..... 3248
705 erGluHisSerLeuAspProGluAspIleArgAspGluLeuLysLysLeu 721
3248 ..... 3248
722 TyrAlaGlnLeuGluIleTyrLysArgLysLysMetIleThrAsnAsnPr 738
3249 .....ATTTCGGAGAAAGAGCCCGG..... 3271
738 oHisLeuGlnLysLysArgCysSerLysLysGlyLeuGlyArgSerIleM 755
3271 ..... 3271
755 etArgArgIleThrGluIleProGluThrValSerArgGlnCysSerLys 771
3272 .....GCAGGAATGCCAAGAA 3289
772 GluAspLysGluGlyAlaAspHisGlyThrAlaLysGlyThrAlaLeuI 788
3290 GAG.....GCAGCCAGAT 3303
```

```
788 eArgLysAsnProProGluSerSerGlyAsnThrGlyLysSerLysGluG 805
3304 TCTCGCCACGACCGCATGTGCTCGGCACATCGCAGCTTTGAAAACCC 3353
805 luThrLeuLysAsnArgValPheSerLeuLysLysSerHisSerThrTyr 821
3354 CCACACTGCAGTGAATGTTCTTACGGCAAGTCTGTGTCATGGTCTGNAAC 3403
822 Asp.....HisValArgAspGlnThrGluGluSerSerSerLe 834
3404 CAGGTGGAAGACAGGCGCCCAAGGACAGCACGTGTGGCAGCGCTCTCT 3453
834 uProThr.GluSerGln.GluGluGluThrThrGluAsnSerThrLeuG 850
3454 GTCACGTGAAGACCAACAGAGAGCGGCTGTAAACCAACAGCCGTAATCAA 3503
850 uSerLeu.....SerGlyLysLysLeuThr..... 858
3504 ACCCTCCTACTAAAGATTACCAAGCTCTGCAAGAGCCTGACCTTTTCAG 3553
859 .....GlnLysLeuLysGluAspSerGluAlaGlu 868
3554 ATGCCAGCACCAAGACCTTTTACAATGTGGAAGAGGACAAATACCCCT 3603
869 SerThrGluSerValProLeuValCysLysSerAlaSerAlaHisAsnLe 885
3604 TCTGCTCACTTCAGCCCTCCAGCAGCCCTTCTATGTTGTGTCACCGACG 3653
885 uSerSerGluLysLysThrGlyHisProArgThrSerMetLeuGlnLys 902
3654 CGGCGCACCGTGGCCACACACCTCTGCCACCCCATCTG..... 3696
902 erLeuSerValIleAlaSerAlaLysGluLysThrLeuGlyLeuAlaGly 918
3697 .....ACCGAGAGAGAGACCCCTGTTCCTGGCTGAT 3729
919 LysThrGlnThrAlaGlyValGluGluArgThrLysSerGlnLysProLe 935
3730 TCCGTCTATCCCCAAGGCGTTG.....CCTCCTCCTCT 3761
935 uProLysAspLysGluThrAsnArgAsnHisSerAsnSerAspAsnThrG 952
3762 CCGCAGCAGCAG..... 3774
952 luThrLysAspProAlaProGlnAsnSerAsnProAlaGluProArg 968
3775 .....CCACAGCAGCGCCCTCAGCAGCCCGCCGAG 3807
969 LysProGlnLysSerGlyIleMetLysGlnArgValAsnProThrTh 985
3808 CAGCCCAAG.....TCCCTGATGGACAGCTGCAAGCGCTAGTCAACAA 3851
985 rAlaAsnSerAspLeu.....AsnProGlyT 994
3852 CTTCCGTTCCGGGATTCCAGATTTCCATGCGGTGTCGGCAGGCCCGGGA 3901
994 hrThrGlnMetLysAspAsnPheAspIleGlyValCysProTrpGlu 1010
3902 CACCA.....GGAAC.....AGC 3915
1011 ValTyrAspLeuThrProGlyProValProSerGlu.....SerLysVa 1025
3916 CTGCGTCTCTGTACCGCCCGCCCTCCCGCAACACCTGCAGATGCT 3965
1025 lGlnLysHisValSerIleValAlaSerGluMetGluLysAsnProThrP 1042
3966 GCCCTGCACCTGAGCACCTTCCAGGAGGAGTCCATCTCCCTCCCTGGGG 4015
1042 heSerLeuLysGluLysSerHisHisLysProLysAlaAlaGluValCys 1058
4016 AGCATCGATGATGACAGT..... 4035
1059 GlnGlnSerAsnGlnLysArgIleAspLysAlaGluValCysLeuTrpG 1075
```

```

4036 .....GAGAGATCAAGTCTCGCAGGAGTTCGTGTACGA 4070
1075 userGlnGly.....GlnSerIleLeuGluAspGluLysLeuLeuI 1089
4071 GCGCGAAGGACACCGAGAGATGAATTCGAAGAGAGGAGGACCTGC 4120
1089 leSerLysThrProValLeuProGluArgAlaLysGluGluAsnGlyGly 1105
4121 CCACAGCAGCAAGCTGACCCCTGAGGATTCCTGCGCCCTGACGCTCCT 4170
1106 GlnProArgAlaAlaAsnValCysAlaGlyGlnSerGluGluLeuProPr 1122
4171 TCTCTCTCCGAGANTCCGTGGCTCTGGCAGCTCAGTCCCGCAGTCCCC 4220
1122 o.....LysAlaValAlaSerLysThrGluValGluAsnGlnI 1137
4221 CGTATCTGAGTGGTCTCTGTCACCCCTCCAAATGTAACTACGCTCTG 4270
1137 leGlyHisGlnGluLysLysThrSerSer 1147
4271 TCATCTCTCAGGAGCTACAAAGCAAGCTCTTCC 4302

```

seq_name: /SIDS1/gcdata/geneseq/geneseq/NAL1992.DAT:AAQ25812

seq_documentation_block:

ID_AAQ25812 standard; cDNA: 4300 BP.

XX AC AAQ25812;
 XX DT 04-JAN-1993 (first entry)
 XX DE Clone 45-A.
 XX KW GLU-G-R; Xenopus oocyte; cerebellum sublibrary; voltage clamp assay;
 KW neurotransmitter; glutamate; neuronal excitation; ss.
 XX OS Synthetic.

XX FH Key
 XX CD 377..3676
 XX FT /*tag= a
 XX FT /label= Subtype_1a

XX WO9210583-A.
 XX 25-JUN-1992.
 XX PF 12-DEC-1991; 91WO-US09422.
 XX PR 12-DEC-1990; 90US-0626806.
 XX PR 30-JAN-1991; 91US-0648481.
 XX PR 18-MAR-1991; 91US-0672007.
 XX (UNIW) UNIV WASHINGTON.
 XX (ZYMO) ZYMOGENETICS INC.
 XX PI Almers W, Hagen FS, Houamed KM, Mulvihill ER;
 XX DR WPI: 1992-234638/28.
 XX DR P-PSDB; AAR25080.
 XX Mammalian G protein-coupled glutamate receptors - activate
 PT phospholipase C or stimulate inositol phospholipid metabolism,
 PT for use in diagnosis and identification of receptor agonists and
 PT antagonists
 XX PS Disclosure; Fig 5; 144pp; English.

XX The sequence given encodes mammalian G protien-coupled glutamate
 CC receptor (GLU-G-R) subtype 1a. The clone was isolated from Xenopus
 CC oocyte cultures which had been injected with RNA extracted from a
 CC cerebellum sublibrary. Expression of GLU-G-R activity was assessed by

CC voltage clamp assay. Plasmid DNA encoding GLU-G-R activity was
 CC isolated by replica-plating experiments. The protein encoded by clone
 CC 45-A was found to act as a receptor for the neurotransmitter glutamate,
 CC and may play a critical role in the control of neurons. GLU-G-R and
 CC antibodies may be used to identify agonists and antagonists of
 CC GLU-G-R-mediated neuronal excitation and in diagnosis.

XX
 SQ Sequence 4300 BP; 1038 A; 1191 C; 1148 G; 923 T; 0 other;

alignment_scores:
 Quality: 231.00 Length: 1237
 Ratio: 0.485 Gaps: 50
 Percent Similarity: 38.480 Percent Identity: 19.078

alignment_block:

US-09-775-181-2 x AAQ25812 ..

Align seg 1/1 to: AAQ25812 from: 1 to: 4300

```

153 AspLeuAspTyrTyrGlnAlaLeuValTrpSerLeuLeuGluGluPr 169
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
853 GATCGGCCCTGGCTCCAGCTCTGTGGCCATTCAGTCCAGATCTTCTCC 902
| | | | | | | | | | | | | | | | | | | | | | | | | | |
169 oSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeu..... 183
| | | | | | | | | | | | | | | | | | | | | | | | | | |
903 A.....GCTGTTCGACATCCACACATCCCTATTCTGCC 937
| | | | | | | | | | | | | | | | | | | | | | | | | | |
184 .....SerAlaProAlaProGlnValPheLeuGlnAlaThrArgGlu.... 197
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
938 ACAAGCATAGACCTGAGTGACAAAACCTTTGTACAAATACTTCCTGAGGT 987
| | | | | | | | | | | | | | | | | | | | | | | | | | |
198 .....GluSerArgIle 201
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
988 GGTCCCTTCTCACACTTTCAGGCAAGGGCGATGCTCCGACATAGTCAAGC 1037
201 eLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrL 218
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1038 GTTACAACCTGGACCTATGTCTCAGCAGTCCACACAGAGGAA..... 1080
218 euGluThrGluTrpPheHisGlyLeuArgArgLysTrp..... 230
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1081 .....TTACGGCAGAGTGGAAATGGATGCTTT 1107
231 .....ArgProHisLeuHisArgArg...GlyPr 239
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1108 CAAAGAACTGGCTGCCAGAGAGGCT...CTGCATCCACACTCGGACA 1154
239 oAsn..... 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1155 AAATCTACAGCAATGCTGGCGAGAAGAGCTTTGACCGGCTCCTCGTAAA 1204
241 .....GlnGlyProArgGlyLeuGlyHisSerTrpArg 251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1205 CTCGGGAGCGGCTTCCCAAGGC...CAGGGTTGTGGTCTGCTTCTGCGA 1251
252 ArgLysAspGlyLeuGlyGlyAspLysSerHis..PheLysTrp..... 265
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1252 GGGCATGACAGTGGGGCTTACTCAGTCCCATGCGCCGCTGGCGCTCG 1301
266 .....SerProProTyrLeuGluCysGluAsnGlySer..... 276
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1302 TGGCGAGTTCTTCTCATTGCAAGTGTATGATGGCAGACAGATGAA 1351
277 .....TyrLysPro..... 279
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1352 GTCATCGAAGGCTATAGGTGCAAGCAACGAGGATTCACAATAAGCT 1401
280 .....Gly 280
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1402 TCAGTCTCCAGAGCTCAGGTCAATTGATGACTACTTCTCGAAGCTGAGGC 1451
281 TrpLeuValThr.....LeuSe 286
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```
1452 TGGACACACACAAAGAAATCCTTGGTTCCTGAGTTCTGCGCAACATCGC ..... 1501
1234 5678 910111213141516171819202122232425262728293031323334353637383940414243444546474849505152535455565758596061626364656667686970717273747576777879808182838485868788899091929394959697989900
286 rSerAlaIleTyrGlyLeuGln.....ProAsnLeu... 296
1502 TTCAGTGTCCCTACCTGGACACCTCTTGGAAAAACCCAACTTTAAGAA 1551
297 .....ValProGluPheArgGlyValMetLysValAspIleAsnLeuGln 311
1552 AGTGTGCAGGAAATGAAGCTTGGAGAAACTATGTCCAGGACAGCA 1601
312 LysValAspIleAspGlnCysSerSerAspGlyTrpPheSerGlyThrHi 328
1602 AAATGGGATTTGTCATCAATGCCATCTATGCCATGGCACATGGGCTCGAC 1651
328 sLysCys..His.....LeuAsnAsnSerGluCysMe 338
1652 AACATGCACCATCTCTGTGTCCGGCCATCTGGGCCCTGTGTGATGCTAT 1701
338 t...ProIleLysGly.....LeuGlyPheValLeu..... 347
1702 GAAACCAATTGATGGGACGAGTCCCTGGATTTCTCATCAAAATCCCTCTT 1751
347 ..... 347
1752 TTGTCCGAGTGTCTGGACAGAGGTGTGGTTTCGATGAGAGGGGGATGCT 1801
348 ...GlyAlaTyrGluCysIle..... 353
1802 CCCGGAAGGTATGACATATCAATCTCCAGTACACACAAGCTAATCCCTA 1851
354 .....CysLysAlaGlyPheTyrHisProGlyValLeuProValAsnA 368
1852 TGACTATGTCCACGTGGGACCTGGCATCTGAAGAGAGTGCTGAATATTGAT 1901
368 snPheArgArgArgGlyProAspGlnHisIleSerGlySerThrLysAsp 384
1902 ATTACAAAATCCAG.....ATGAACAAAAGCGGAATGGTACGATCT 1942
385 ValSerGluGlu..... 388
1943 GTGTGCAGTCAGCCTCTTAAAGGGTCAGATTAAAGTCATACGGAAGG 1992
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
1993 AGAAGTGAGCTGTGCTGTGGATCTGACGGCTGCAAAAGAAATGAGTTG 2042
399 .....CysProPheCys..... 402
2043 TGCAGGACGAGTTCACCTGCAGACCTGTGACCTGGGGTGGTGGCCCAAC 2092
403 AlaAspAspSerProCysPheValGlnGluAspLysTyrLeuArg..... 417
2093 GCAGAGCTCACAGCTGTGAGCCCATTCCTGCTCGGTATCTTGAGTGGAG 2142
418 .....LeuAlaIleLeuSerPheGlnGlyLeuCysMetLeuL 430
2143 TGACATAGAATCATCATAGCCATCGCCTTTCTTGCTGGGCATCCTCG 2192
430 euAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSer 446
2193 TGACGCTGTTTGCACCTCATCTTCGTTCTGTACCGGGACACACCCGTG 2242
447 IleArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLe 463
2243 GTCAATCTCCAGTAGGAGCTCTGCTATATCATTTCTGCTGGCTGATTTT 2292
463 uLeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheA 480
2293 CCTCGGCTATGTGTGCCCTTTCACCTCATCGCAAAACCTACTACCACAT 2342
480 rgCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyr 496
1234 5678 910111213141516171819202122232425262728293031323334353637383940414243444546474849505152535455565758596061626364656667686970717273747576777879808182838485868788899091929394959697989900
```

```

763 GluThrValSerArgGlnCysSerLysGluAspLysGluGlyAlaAspHI 779
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3128 AAGACCAACGAGCGCCGTAAACAA..... 3154
779 sGlyThrAlaLysGlyThrAlaLeuIleArgLysAsnProProGluSerS 796
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3155 .....ACAGCGGTAAACAA..... 3169
796 erGlyAsnThrGlyLysSerLysGluThrLeuLysAsnArgValPhe 812
3169 ..... 3169
813 SerLeuLysLysSerHisSerThrThrAspHisValArgAspGlnThrG 829
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3170 CCCCTCACTAAAGTTAC..... 3187
829 uGluSerSerLeuProThrGluSerGlnGluGluThrThrGluA 846
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3188 .....CAAGGCTCTGGCA 3200
846 snSerThrLeuGluSerLeuSerGlyLysLysLeuThrClnLysLeuLys 862
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3201 ACAGCCTGACCTTTTCAGATGCCAGCACCAAGACCCCTTTACAAATGTGGAA 3250
863 GluAspSerGluAlaGluSerThrGluSerValProLeuValCysLysSe 879
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3251 GAAGAGGACAATACCCCTTCTCCTCACTACCCCTCCAGCAGCCCTTC 3300
879 rAlaSerAlaHisAsnLeuSerSerGluLysLysThrGlyHisProArgT 896
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3301 TATGTTGGTGCACCGAGCGGCCACCGCTGGCCACACACACCTCTGC 3350
896 hrSerMetLeuGlnLysSerLeuSerValIleAlaSerAlaLysLys 912
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3351 CACCCCATCTG.....ACCGCAAGAGAGACC 3376
913 ThrLeuGlyLeuAlaGlyLysThrGlnThrAlaGlyValGluGluArgTh 929
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3377 CCCCTGTTCTGGCTGATTCGCTCATCCCAAGGCGTTG..... 3415
929 rLysSerGlnLysProLeuProLysAspLysGluThrAsnArgAsnHisS 946
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3416 .....CCTCTCTCTCTCCCGCAGCAGCAG..... 3439
946 erAsnSerAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsn 962
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3440 .....CCACAGCAGCGCGCC 3454
963 ProAlaGluGluProArgLysProGlnLysSerGlyIleMetLysGlnG 979
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3455 CCTCAGCAGCCCCCGCAGCAGCCCAAG.....TCCCTGATGGACCACT 3498
979 nArgValAsnProThrThrAlaAsnSerAspLeu..... 990
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3499 GCAAGGCGTAGTCACCAACTTCGGTTCGGGATTCAGATTTCATCGCG 3548
991 .....AsnProGlyThrThrGlnMetLysAspAsnPhe 1001
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3549 TGCTGGAGCGCGCGGACACAGGAAACAG..... 3580
1002 AspIleGlyGluValCysProThrGluValTyAspLeuThrProGlyPr 1018
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3581 .....CTGGCCTCTCTGTACCCGCCGCC 3603
1018 oValProSerGlu.....SerLysValGlnLysHisValSerIleValA 1033
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3604 GCCTCCGCGCAACACCTCGAGATGTCGCCCTGCACCTGACGACCTTCC 3653
1033 IaSerGluMetGluLysAsnProThrPheSerLeuLysGluLysSerHis 1049
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3654 AGGAGGATCCATCTCCCTCTCTGGGAGGACATCGATGATGACAGT... 3700

```

```

1050 HisLysProLysAlaAlaGluValCysGlnGlnSerAsnGlnLysArgI 1066
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3701 .....GAGAGATT 3708
1066 easLysAlaGluValCysLeuTrpGluSerGlnGly.....GlnS 1080
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3709 CAAGCTCTCTGTCAGGAGTTGCTGTACGAGCGCGAAGGACACCGAAGAG 3758
1080 erIleLeuGluAspGluLysLeuIleSerLysThrProValLeuPro 1096
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3759 ATCAATTGGAAGAGGAGGAGGACCTGCCACAGCAGCAGCTGACCCCT 3808
1097 GluArgAlaLysGluGluAsnGlyGlnProArgAlaAlaAsnValCy 1113
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3809 GAGGATTCTCTCTGCTGACGCTCCTCTCTCCGAGATTCCGTGGC 3858
1113 salGlyGlnSerGluGluLeuProPro.....LysAlaValAlaSerL 1128
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3859 CTCTGGCAGCTCAGTCCCGAGTTCCCGCTATCTGAGTCGTCTCTGCA 3908
1128 ysThrGluAsnGluAsnLeuAsnGlnIleGlyHisClnGlnLysLysThr 1144
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3909 CCCCTCCAAATGTAACCTACGCTCTCTGCAATCTCAGGAGACTACAAGCA 3958
1145 SerSerSer 1147
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3959 AGCTCTTCC 3967

```

seq_name: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ31058

seq_documentation_block:

ID AAZ31058 standard; cDNA; 3831 BP.

XX AC AAZ31058;

XX DT 07-JAN-2000 (first entry)

XX DE phCar/hmGluR2*Gq15 fusion construct cDNA sequence.

XX KW G-protein fusion receptor; CaR; calcium receptor; GluR; head injury;
 KW metabotropic glutamate receptor; GABABR; chimeric receptor; stroke;
 KW gamma-aminobutyric acid receptor; allosteric modulator; antagonist;
 KW spinal cord injury; epilepsy; ischaemia; hypoglycaemia; anoxia;
 KW Alzheimer's disease; hyperparathyroidism; osteoporosis; depression;
 KW cognitive disorder; ss.

XX OS Homo sapiens.

XX PN WO9951641-A1.

XX PD 14-OCT-1999.

XX PF 02-APR-1999; 99WO-US07333.

XX PR 03-APR-1998; 98US-0080671.

XX PA (NPSP-) NPS PHARM INC.

XX PI Stormann TM, Hammerland LG, Storjohann LL, Busby JG, Garrett JE;

XX PI Simin RT;

XX DR WPI; 1999-610995/52.

XX DR P-PSDB; AAY49127.

XX PT New G-protein fusion receptors and chimeras containing domains from
 PT different receptors, used to screen for modulators, potentially useful
 PT e.g. for treating or preventing stroke or Alzheimer's disease.

XX PS Example 1; Fig 11; 255pp; English.

XX CC The invention relates to G-protein fusion receptors (I) comprising:

XX CC (1) in the N to C direction, extracellular (ECD), transmembrane (TMD) and
 XX CC intracellular (ICD) domains, each chosen independently from a CaR

CC (calcium receptor), GluR (metabotropic glutamate receptor) and GABABR
 CC (gamma-aminobutyric acid receptor); (2) an optional linker attached to
 CC the C-terminus of ICD; and (3) a G-protein (GP) linked to ICD or the
 CC linker. (1), and recombinant chimeric receptors (CR) without the GP
 CC component, are used to assess function of the various domains and to
 CC identify compounds (e.g. allosteric modulators or antagonists) that act
 CC on these domains. The modulators are potentially useful for treating or
 CC preventing diseases associated with the receptors, e.g. stroke, head or
 CC spinal cord injury, epilepsy, ischemia, hypoglycemia, anoxia,
 CC Alzheimer's disease, hyperparathyroidism, osteoporosis, cognitive
 CC disorders and depression. Nucleic acid (II) that encodes (I) is used:
 CC (1) for recombinant production of corresponding proteins; and (2) to
 CC produce cells used in screening for modulators. Use of CAR and mGluR
 CC domains allows presentation of GABABR domains, to a binding agent, in a
 CC form more like the natural domain structure compared with use of
 CC incomplete receptors, lacking one or more domains. By shuffling different
 CC domains, agents can be identified that affect particular domains of a
 CC receptor.

xx
 SQ Sequence 3831 BP; 917 A; 1025 C; 998 G; 891 T; 0 other;

alignment_scores:

Quality: 217.50 Length: 869
 Ratio: 0.583 Gaps: 38
 Percent Similarity: 42.923 Percent Identity: 20.253

alignment_block:

us-09-775-181-2 x AAZ31058 ..

Align seg 1/1 to: AAZ31058 from: 1 to: 3831

```

75 AlaGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspse 91
   :::::.....:::
394 TCAGAGCACATTCCTAGGATTGCTGTGGTGGGAGCAACTGGCTCAGG 443
   91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
   444 CGTCTCCACGGCAGTGGCAAAATCGTGGGGCTCTTACATCATT..... 486
108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
   ||| ::::: ||| ||| |||
487 .....CCCCAGTCAGTTATGCCCTCCACG.....AGA 516
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG 141
   ||| ::::: ||| ||| ||| |||
517 CTCCTCAGCAACAAGATCAATCAAGTCTTCTCGGAACC...ATCCC 563
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAspLeuTrpTyrG 158
   ::::: ||| ||| ||| ||| ||| |||
564 CAATGATGAGCACCGCCACTGCCATGGCAGACATCATCGATATTTC 613
158 lnAlaLeuValTrpSerLeuLeuGluGlyGluProSerIleSer..... 172
   ::::: ||| ||| ||| ||| ||| |||
614 GCTGGAAGTGGTGGGCACAAATTCGACCTGATGACACATATGGCGGCCG 663
173 .....ArgAlaAlaIleThrPheSerT 180
   ||| ||| ||| ||| ||| |||
664 GGGATTGAGAAATTCGAGAGGAGTGGAGAAAGGGATATCTGCATCGA 713
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
   ||| ||| ||| ||| ||| ||| |||
714 CTTCACTGAAGTATCTCCAGTACTGTGATGAGAGAGATCCAGCATG 763
196 .....ArgGluGluSerArgIleLeuLeuG 204
   ||| ||| ||| ||| ||| ||| |||
764 TGGTAGAGGTGATTCAAATTCACGCCCAAGTCACTCGTGGTTTCTCC 813
204 lnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuLeuThr 220
   ||| ||| ||| ||| ||| ||| |||
814 AGTGGCCAGATCTTGAGCCCTCATCAAGGAGATTGTCGGCGCA... 860
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArg 237

```

```

861 .....TATCAGGGGCAAGATCTGGCTGGCCAGGAGGCTGGGCCAGCT 904
   ::::: ||| ||| |||
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
   ||| ::::: ||| ||| ||| |||
905 CCTCCCTGATCGCCATCCCTCAGTACTTCACGCTGGTGGCGCACCAT 954
   ||| ||| ||| ||| ||| |||
250 rpArgArgLysAspGlyLeuGlyClyAspLysSerHisPheLysTrpSer 266
   ||| ||| ::::: ||| ||| |||
955 GGATTCGCTCTGAAGG.....TGGGCA 977
267 ProProTyrLeuGlu.....Cy 272
   ||| ||| ||| ||| |||
978 GATCCCAAGGCTTCGGGAATTCCTGAAGAAGTCCATCCCAAGTCTG 1027
272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
   ||| ||| ||| ||| ||| |||
1028 TCCACAATGCTTTTGGCAAGGAGTTTGGGAAGAACATTTAACTGCCAC 1077
285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
   ||| ||| ||| ||| ||| |||
1078 CTCCAAGAGGTGCAAAAGGACCTTTTACCTGTGGACACCTTTCTGAGAGG 1127
295 .....AsnLeuValProGluPheArgG 302
   ||| ||| ||| ||| ||| |||
1128 TCAGGAAGAAGTGGCGACAGGTTTACCAACAGCTCGACAGCCTTCCGAC 1177
302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
   ||| ||| ||| ||| ||| ||| |||
1178 CCCTCTGTACAGGGGATGAGAACATCAGCAGTGTGAGAGCCCTTACATA 1227
313 ..... 313
1228 GATTACACGCATTTACGGATATCCTACAATGTGTACTTAGCAGTCTACTC 1277
314 .....AspIleAspGlnCys..... 318
1278 CATTGCCCGCCCTTGCAGATATATATACCTGTACCTGGGAGAGGGC 1327
319 .....SerSerAspGlyTrpPhe 324
   ||| ||| ||| ||| ||| |||
1328 TCITCACCANTGCTCCTGTGCACACATCAAGAAAGTTGAGCGCTGGCAG 1377
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
   ||| ||| ||| ||| ||| ||| |||
1378 GTCTGAAGCACCTACGGCATCTAAACTTTACAAACAATATGGGGAGCA 1427
336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
   ||| ||| ||| ||| ||| ||| |||
1428 GGTGACCTTTGATGAGTGTGTGACCTGGTGGGGAACATTTCCATCATCA 1477
348 .....Gly 348
   ||| ||| ||| ||| ||| |||
1478 ACTGGCACCTCTCCCCAGAGGATGGCTCCATCGTGTTTAAGGAAGTCGG 1527
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProG 362
   ||| ||| ||| ||| ||| ||| |||
1528 TATTACAACGTCTATGCCAAGAAGGAGAAAGACTCTTTCATCAACAGGA 1577
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis.... 377
   ||| ||| ||| ||| ||| ||| ||| |||
1578 GAAATCCTGTGGAGTGGTGTCCAGGAGGAGTGCCTTCTCCAACATGCA 1627
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
   ||| ||| ||| ||| ||| ||| ||| |||
1628 GCCCAGACTCCTGGCAGGACCAAGGAAAGGATCATTTAGGGGGAGCC 1677
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
   ||| ||| ||| ||| ||| ||| ||| |||
1678 ACCTGCTGCTTGTAGTGTGTGGAGTGTCTGTATGGGAGTATAGTGATGA 1727
399 .....CysProPheCysAlaAspAsp..... 405
   ||| ||| ||| ||| ||| |||

```

1728 GACAGATGCCAGTGGCTGTACAAAGTGGCCAGATGACTTCTGCTCCAATG 1777
406SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
::: ||||| ::::: |||||
1778 AGAACACACACCTCTGCTTCAAGATGCCCCAGGAGTACATCCGCTGGGCG 1827
418LeuAlaIleLeuSerPheGlnGlyLeuCy 427
::: ||||| ::::: |||||
1828 GATGCTGGGCTGTGGGAGCTGTACCATCGCTGCCTCGGTGGCTGGC 1877
427 sMetLeuLeuAspPheValSerMetLeuValValTyrHisPheArgLysA 444
1878 CACCCTG.....TTTGTCTGGGTGCTTTGTGGGCAC...AATGCCA 1918
444 laLysSerIleArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPhe 460
::: ||||| |||||
1919 CACCAGTGTCAAGCCCTCAGGTGGGAGCTGTGTACATCTGTGTGGGT 1968
461 GlySerLeuLeuTyrPheProValIleLeuTyrPheGluProSe 477
||| ::::: ||| ::::: |||||
1969 GGTGTCTCTCTGTCTGTCTGTACATGACCTTCATCTTCATTGGCAAGCCATC 2018
477 rThrPheArgCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaT 494
2019 CACGGCAGTGTGTACCTACGGCGTCTGGTTGGGCACGTGCTCTCTCTG 2068
494 hrValTyrGlyThrValThrLeuLysLeuHisArgValLeuLysValPhe 510
::: ||||| ::::: |||||
2069 TCTGTCTACTCAGCCCTGTCTACCAAGACCAACCGCATTCGACGCATCTTC 2118
511LeuSerArgThrAlaGlnArgIleProTyrMetThr...GlyCl 524
::: ||||| |||||
2119 GGTGGGCGCGGAGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2168
524 yArgValMetArgMetLeuAlaValIle.....LeuLeuValV 537
2169 ACAGTGGCCATCTCCGTGGCACTTATCTCGGGCCAGCTGTCTCACTGCTGG 2218
537 alPheTrpPheLeuIle.....GlyTrpThrSerSerValCysGln 550
||| ||||| ::::: |||||
2219 TCGCCTGTGTGTGTGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2268
551 AsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAsp... 566
::: ||||| ::::: |||||
2269 GAACGGCGGAGGT 2318
567HisLeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrM 581
||| ||||| |||||
2319 GTTGGGCTCGTGGCTACAATGTGCTCTCATC..... 2352
581 etThrAlaValAlaGluPheLeuPheLeuTrpGlyValTyrLeuCys 597
2353GGCCTGTGCAGCTTAT...GCC 2373
598 TyrAlaValArgThrValProSerAlaPheHisGluProArgTyrMetAl 614
::: ||||| |||||
2374 TCAATACCTCGCAAGTGTCCGGAACCTTCAACGAGGCGCAAGTTCATTGG 2423
614 aValAlaValHisAsnGluLeuIleIleSerAlaIlePheHisThrIleA 631
::: ||||| |||||
2424 CTTACCAATGACACCACTGCATCATCTGTGTCATGTGTGTCCTCTCTCT 2473
631 rPheValLeuAlaSer.....ArgLeuGlnSerAspTrpMetLeuMet 645
::: ||||| |||||
2474 TCTATGTACCTCCAGTACTACCGGGGTACAGACCAACCACTGTGCTGG 2523
646 LeuTyrPheAlaHisThrHisLeuValThrValThrIleGlyLeuLe 662
::: ||||| ::::: |||||
2524TCAGTACGCTCAGCGGCTCGGTGTGCTGTGCTGTGCTGTGCT 2561
662 uLeuIleProLysPhe..... 667
2562 CTTTGGCCCAAGCGACATCATCTCTTCCAGCGCGGAGAACGATGG 2611

668 ..SerHisSerSerAsnAsnProArgAspIleAlaThrGluAlaTyr 683
||||| ::::: |||||
2612 TTAGCCACCGGGCACCCACAGCGCTTTGGCAGTGTCTGCCAGGCGC 2661
684 GluAspGluLeuAspMetGlyArgSerGlySerTyr..... 695
::: ||||| ::::: |||||
2662 AGCTCAGCCTTGGCAAGGCTGTGGCTCCAGTTGTGCCACCTGTGTTG 2711
696LeuA 697
2712 CAATGGCGGTGAGTGGGACTCGACACGTCATCGCTTATGACTCTGG 2761
697 snSerSerIleAsnSerAlaTrpSerGluHisSerLeuAspProGluAsp 713
::: ||||| ::::: |||||
2762 AGTCATCATGCGTGTGCTGTGAGCGAGGAGGCCAAGAACCGCGCGG 2811
714 IleArgAspGluLeuLysLysLeuTyrAlaGlnLeuGluIleTyrLysAr 730
||||| ::::: |||||
2812 ATCAACGACGAGATCGAGCG.....CAGCTCCGAGGAGCAAGCG 2852
730 g 730
2853 G 2853
seq_name: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ10801
seq_documentation_block:
ID AAZ10801 standard; DNA: 3591 BP.
XX AC AAZ10801;
XX DT 09-NOV-1999 (first entry)
XX DE Mouse metabotropic glutamate receptor 5 (mGluR5) gene.
XX KW Metabotropic glutamate receptor; mGluR5; central nervous system; CNS;
KW depressant; anaesthesia; phospholipase C; intracellular calcium; ss.
XX OS Mus sp.
XX FH Key Location/Qualifiers
FT CDS 1..3591
FT FT /*tag= a
FT FT /product= mGluR5
FT FT /note= "metabotropic glutamate receptor 5"
XX PN WC9938975-A2.
XX PD 05-AUG-1999.
XX PE 29-JAN-1999; 99WO-US02033.
XX PR 30-JAN-1998; 98US-0073190.
XX PA (UYTE-) UNIV TECHNOLOGY CORP.
XX PI Johnson TE, Rikke BA, Sikela JM, Simpson VJ;
XX WPI: 1999-494093/41.
XX DR P-PSDB; AAY28562.
XX CNS depressant sensitivity associated a YAC containing murine
PT tyrosinase and metabotropic glutamate receptor 5 genes, useful for
PT identifying CNS depressant modulators
XX Claim 21; Fig 14; 140pp; English.
XX This sequence is the mouse metabotropic glutamate receptor 5 gene
CC (mGluR5). mGluRs are G protein-coupled receptors capable of activating a
CC variety of second messenger systems. mGluR5 belongs to mGluR Group I. The
CC binding of an agonist to group I receptors results in the activation of
CC phospholipase C causing the mobilisation of intracellular calcium. The

1605 TTGTAAGGAGAA TGAGTATGTTT TTGATGAGTACACCTGCAAGGCGTGCC 1654

```

662 LeuLeuIleProLys.....PheSerHisSerSerAsnAsnPr 674
||||:|||||
2446 ATGTTTGTCCGAAGGTGTACATCATCTAGCCAAACCGGAGAGAAATGT 2495
674 oArgAspAspIleAlaThrGlu...AlaTyrGluAspGluLeuAspMetG 690
||||:|||||
2496 GCGCAGCGCTTCACAACTCTACAGTGTGTGCGCATGCGACGTAGGAGATG 2545
690 lYargSerGlySerTyrLeu.....AsnSerSerIleAsnSerAlaTrp 704
||||:|||||
2546 GCAAGTCATCATCCGCTGCCAGCAGATCCAGCAGCCTAGTCAACCTGTGG 2595
705 SerGluHisSerLeuAspProGluAspIleArgAspGluLeuLysLysLe 721
||||:|||||
2596 AAGAGGAGGGGTCTGTGGGAACCTTAAGGTACAAACAGCAGGAGACT 2645
721 u.....TyrAlaGlnLeuGluIleTyrLysArgLysLysMetIleT 735
||||:|||||
2646 GGCACGACACAGTCGGAATAGAGTGTTCACCCCAANGGAGATATGG 2695
735 hrAsnAsnProHisLeuGlnLysLysArgCysSerLysLysGlyLeu... 750
||||:|||||
2696 GGAATGCTGGAGAGCAACAATGAGCAGCTCCAATGGAATAATCCGTGACT 2745
751 .....GlyArgSerIleMetArgArg.. 757
||||:|||||
2746 TGGGCCCAAGATGAGAAGAGCACCCCGGGGCGACACCTGTGGCAGGACT 2795
758 .....IleT 759
2796 GTCTGTCCATATCACAGAAGAGAGAACCCCAACCCAGACAGCAGTCATCA 2845
759 hrGluIleProGluThrValSerArgGlnCysSerLysGluAspLysGlu 775
||||:|||||
2846 AACCTCTCCCAAGACACAGAG.....AGCCGCGGCGAGGGTGCA 2886
776 GlyAlaAspHisGlyThrAlaLysGly..... 784
||||:|||||
2887 GGGGCGAGGTGGTGGTCTGGCCCGCGTGCGTGGTGGTAGGCGCAGG 2936
785 ....ThrAlaLeuIleArgLysAsnProProGluSerSerGlyAsnThrG 800
||||:|||||
2937 ATGCACAGCGCAGCGGGCCAGCCAGCCAGCCAGCCGCGCCCAAGCGG. 2985
800 lYlysSerLysGluGluThrLeuLysAsnArgValPheSerLeuLysLys 816
2985 ..... 2985
817 serHisSerThrTyrAspHisValArgAspGlnThrGluGluSerSerSe 833
||||:|||||
2986 .....CTTTATGATGTCGACAGAGCGCAGAGCGG 3014
833 rLeuProThrGluSerGlnGluGluThrThrGluAsnSerThrLeuG 850
||||:|||||
3015 CTTCACAGCGGTGCAGCGCGCTGCGCATCGCCATCATGACGCTGCA 3064
850 luSerLeuSerGlyLysLysLeuThrGlnLysLeuLysGluAspSerGlu 866
||||:|||||
3065 GCCACCTGCGCAGC.....TCGCGC 3084
867 AlaGluSerThrGluSerValProLeuValCysLysSerAlaSerAlaHi 883
||||:|||||
3085 GCGCCACAGCAGCAGCAGCGCGCTGCGTGCACCTCGGACCGCTGCGACG 3134
883 sAsnLeuSerSerGluLysLysThrGlyHisProArgThrSerMetLeuG 900
||||:|||||
3135 CAGCAGCTCATCCCGGGCTGCTCATGAGCAGATTAGCAGCGTGGTGA 3184
900 lNlysSerLeuSerValIleAlaSerAlaLysGluLysThrLeuGlyLeu 916
||||:|||||
3185 CCGCGTTCCACGCCCAACATCACCGCAGCTCAACTCCATGATGTGTCCACC 3234
917 AlaGly..... 918

```

seq_name: /sids1/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ80419

seq_documentation_block:

ID: AAQ80419 standard; cdna; 4085 BP.

XX AC AAQ80419;

XX AC

DT 21-JUL-1995 (first entry)

XX DE Human mGluR5a cDNA.

XX KW Metabotropic glutamate receptor; mGluR5; mGluR5a; hippocampus;

XX KW cerebellum; pCMV-T7-3; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 370..3912

FT /*tag= a

FT /product= human mGluR5a

XX PN W09429449-A.

XX PD 22-DEC-1994.

XX PF 03-JUN-1994; 94WO-US06273.

XX PR 04-JUN-1993; 93US-0072574.

XX PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.

XX PI Daggett L, Ellis SB, Hess SD, Johnson EC, Liaw C;

XX PI Pontsler A;

XX DR WPI; 1995-036478/05.

XX DR P-PSDB; AAR64253.

XX PT New DNA encoding human metabotropic glutamate receptor

XX PT sub: type(s) - and related proteins, probes, RNA, transformed

XX PT cells and antibodies, useful in treatment, diagnosis and

XX PT identification of specific modulators

PS Claim 16; Page 90-95; 125pp; English.
 XX A human hippocampus cDNA library was screened with DNA encoding
 CC the rat mGluR1 receptor. An isolated clone was used to screen
 CC human cerebellum cDNA libraries to isolate additional clones. 3
 CC 3 splice variants of the human mGluR5 transcript, mGluR5a (given
 CC in AA080419), mGluR5b (AA080520) and mGluR5c (AA080421), encoding the
 CC proteins given in AA04253-55, respectively, were constructed by
 CC ligating different overlapping clones. Coding sequences are linked
 CC to regulatory elements of pCMV-T7-3 for expression in mammalian
 CC cells.
 XX
 SQ Sequence 4085 BP; 971 A; 1103 C; 1102 G; 909 T; 0 other;

alignment_scores:
 Quality: 213.00 Length: 1084
 Ratio: 0.457 Gaps: 42
 Percent Similarity: 42.989 Percent Identity: 18.358

alignment_block:

US-09-775-181-2 x AA080419 ..

Align seg 1/1 to: AA080419 from: 1 to: 4085

```

219 GluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHi 235
      ::::::::::::::|
1366 CATGTCAGAGTGGTTTGATGATTATTCGAAGCTCCGCCGAGAAACAAA 1415
      ::::::::::::::|
235 SARArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrpArgA 252
      ::::::::::::::|
1416 CCACCGAANCCCT.....TGGTTTCAAGAAATTTGGCAGC 1450
      ::::::::::::::|
252 rGLysAspGlyLeuGlyAspLysSerHisPheLysTrpSerProPro 268
      ::::::::::::::|
1451 ATCGT.....TTTCAGTCCGACTGGAA 1473
      ::::::::::::::|
269 TyrLeuGluCysGluAsnGlySerTyr..... 277
      ::::::::::::::|
1474 GCGTTTCCACAGGAGACAGCAACAAATACAAAGACTTGCAATAGTTCTCT 1523
      ::::::::::::::|
278 .....LysProGlyTrpLeuVal. 283
      ::::::::::::::|
1524 GACTCTGAAACACATCATGTTCAGGATTCCAAAATGGGATTTGGTATCA 1573
      ::::::::::::::|
284 ...ThrLeuSerSerAlaIleTyrGlyLeu.....GlnProAsnLeu 296
      ::::::::::::::|
1574 ACGCATCTATTTCGATGGCCCTATGGGCTCCACACATGCATGTCCTC 1623
      ::::::::::::::|
297 ValProGluPheArgGlyVal.....Me 304
      ::::::::::::::|
1624 TGCCCGAGCTATGCAGGACTCTGTGATGCCATGAAGCCAAATTGATGGACG 1673
      ::::::::::::::|
304 tLysValAspIleAsnLeuGlnLysValAspIleAspGlnCysSerSera 321
      ::::::::::::::|
1674 GAATCTTTGGAGTCCCTGATGAACCAACAAATTTACTGGGTTTCTGGAG 1723
      ::::::::::::::|
321 spGlyTrpPheSerGlyThrHisLysCysHisLeuAsnAsnSerGluCys 337
      ::::::::::::::|
1724 ATACGATCTATTTCGATGAGATGGAGACTCTCCAGGAGGATTTGAATA 1773
      ::::::::::::::|
338 MetProIleLysGlyLeuGlyPheValLeuGlyAlaTyrGluCysIleCy 354
      ::::::::::::::|
1774 ATGAATTTCAAGGAATGGGA..... 1794
      ::::::::::::::|
354 sLYsAlaGlyPheYrHisProGlyValLeuProValAsnAsnPheArgA 371
      ::::::::::::::|
1795 .AAAGATTACTTTGATTATATCAACGTTTGGAAAGTTGGACATGGAGAA 1843
      ::::::::::::::|
371 rArgGlyProAspGlnHisIle.....SerGlySerThrLys 383
      ::::::::::::::|
1844 TAAAAATGGATGATGATGAAGTATGGTCCCAAGAAACCAACATCATCAGA 1893
      ::::::::::::::|

```

```

384 AspValSerGluGlu..... 388
      ::::::::::|
1894 TCTGTGTGACGTGAACCATGTGAGAAAGCCAGATCAAGGTGATCCGAAA 1943
      ::::::::::|
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
      ::::::::::|
1944 GGGAGACTCAGCTGTTCTGGACCTGTACACCTTGTAAAGGAGATGAGT 1993
      ::::::::::|
399 .....CysProPheCys..... 402
      ::::::::::|
1994 ATGTCTTTTGATGAGTACACATGCAAGGCATGCCAACTGGGGTCTTGGCCC 2043
      ::::::::::|
403 AlaAspAsp....SerProCysPheValGlnGluAspLysTyrLeuArg... 417
      ::::::::::|
2044 ACTGATGATCTCACAGGTGTGACTTGTATCCAGTACAGTATCTTCGATG 2093
      ::::::::::|
418 .....LeuAlaIleIleSerPheGlnGlyLeuCysMetL 429
      ::::::::::|
2094 GGTGACCCCTGAACCATGTGAGCTGTGGTGTTCCTGCTTGGCCTTGGCCTCC 2143
      ::::::::::|
429 euLeuAspPheValSerMetLeuValTyrHisPheArgLysAlaLys 445
      ::::::::::|
2144 TGCCACCCCTGTTGTACTGTAGTCTTCATCATTTACCGTGATACACCA 2193
      ::::::::::|
446 SerIleArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySe 462
      ::::::::::|
2194 CTAGTCAAGTCTCAAGCAGGAACTCTGCTACATATTATCTTGTCTGGCAT 2243
      ::::::::::|
462 rLeuLeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrP 479
      ::::::::::|
2244 CTGCTGGGCTACTTATGATACCTTCTGCTCATTTGCGAAGCCCAACAGA 2293
      ::::::::::|
479 heArgCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrVal 495
      ::::::::::|
2294 TTTACTGCTACTTCAGAGAAATGGCATTTGCTCTCTCCCGACCATGAGC 2343
      ::::::::::|
496 TyrGlyThrValThrLeuLysLeuHisArgValLeuLysVal..... 509
      ::::::::::|
2344 TACTCAGCCCTTGTAAACAAAGACCAACCGTATTGCAAGGATCTCTGGCNG 2393
      ::::::::::|
510 .....PheLeuSerArgThrA 515
      ::::::::::|
2394 CAGCAAGAAAGATCTGTACCCCAAGCCAGATTCATGATGCCTGTG 2443
      ::::::::::|
515 laGlnArgIleProTyrMetThrGlyArgValMetArgMetLeuAla 531
      ::::::::::|
2444 CCCAG..... 2448
      ::::::::::|
532 ValIleLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerSerVa 548
      ::::::::::|
2449 .....CTAGTGAATGCTTTCATTCATCATA..... 2472
      ::::::::::|
548 lCysGlnAsnLeuGluLysGlnIleSerLeu..... 558
      ::::::::::|
2473 .TGATCCAGTTGGGCATCATCGTTGGCCCTCTTTATATGAGGCTCCTG 2521
      ::::::::::|
559 ..IleGlyGlnGlyLysThrSerAspHisLeuIlePheAsnMetCysLeu 574
      ::::::::::|
2522 ACATAATCATGACTACCCAGCATTCGAGAAGTCTACCTGATCTGTAAAC 2571
      ::::::::::|
575 IleAspArgTrpAspTyrMetThrAlaValAla...GluPheLeuPheLe 590
      ::::::::::|
2572 ACCACCAACCTAGGAGTGTGCATCCACTTGGAAACAATGATGTTGAT 2621
      ::::::::::|
590 uLeuTrpGlyValTyrLeuCysTyrAlaValArgThrValProSerAlap 607
      ::::::::::|
2622 TTTGAGTGCACCTCTATGCGTTCAAGACCAAGAAATGTTCCAGCTAACT 2671
      ::::::::::|
607 heHisGluProArgTyrMetAlaValAlaHisAsnGluLeuIleIle 623
      ::::::::::|
2672 TCCCGGAGGCCAAGTATATCCCTTCACAATGTACAGACCTGCATTATA 2721
      ::::::::::|
624 SerAlaIlePheHisThrIleArgPheValLeuAlaSerArgLeuGlnSe 640
      ::::::::::|

```



```
314 .....AspIleaspGlnCys..... 318
1278 CATTTGCCACGCTTGAAGATATATACCTGCTTACCTGGAGAGGGC 1327
319 .....SerSerAspGlyTyrPhe 324
1328 TCTTCACCAATGCTCTGTGCGACATCATCAGAAAGTTGAGGCGTGGCAG 1377
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
1378 GTCTCTGAAGCACCTACCGCATCTFAAACTTTACAAACAATATGGGGAGCA 1427
336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu.. 347
1428 GGTGACCTTTGATGAGTGTGGTACCTGTGGGGAACCTATTCCATCATCA 1477
348 .....Gly 348
1478 ACTGGCACCTCTCCCGAGGAGTGGCTCCATCGTGTTTAAGGAAGTCGGG 1527
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProG1 362
1528 TATTACAAGCTCTATGCCAAGAAGGAGAAAGACTCTTCATCAACAGGGA 1577
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis.... 377
1578 GAAATCCTGTGGAGTGGTGTCCAGGAGGTGCCCTTCTCCAACTGCA 1627
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
1628 GCCGAGACTGCTGGCAGGACCAGGAAGGATCATTTGAGGGGAGGCC 1677
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
1678 ACCTGCTGCTTTGAGTGTGTGGAGTGTCTCGATGGGGAGTATAGTATGA 1727
399 .....CysProPheCysAlaAsp..... 405
1728 GACAGATGCCAGTGGCTGTAAACAAGTCCAGATGACTTCTGGTCCAATG 1777
406 .....SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
1778 AGNACCACACTCTGCTTCGACTGCCAGGACCCAGGATACATCCGCTGGGCG 1827
418 .....LeuAlaIleIleSerPheGlnCysLeuCy 427
1828 GATGCTGGGCTGTGGGACCTGTCCACCATCGCTGCTCGGTGCCCTGGC 1877
427 sMetLeuLeuAspPheValSerMetLeuValValTyrHisPheArgLysA 444
1878 CACCCCTG.....TTGTGCTGGGTGTCTTTGTGGGCGAC...AATGCCA 1918
444 lAlaSerIleArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPhe 460
1919 CACCAGTGTCAAGGCTCAGTGGGAGCTGTGTACATCTCTGTGGT 1968
461 GlySerLeuLeuLeuTyrPheProValIleLeuTyrPheGluProSe 477
1969 GGTGCTCTCTCTGCTACGTACGTACCTTCATCTTCATTGCCAAGCCATC 2018
477 rThrPheArgCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaT 494
2019 CACGGCAGTGTCTACCTTACGGCGTCTTGGTTGGGCACTGCTCTCTCTG 2068
494 hrValTyrGlyThrValThrLeuLysHisArgValLeuLysValPhe 510
2069 TCTGTACTACCTGTCTCACACAGACCAACCGCATTCACGACATCTTC 2118
511 .....LeuSerArgThrAlaGlnArgIleProTyrMetThr...GlyG1 524
2119 GGTGGGCGCGGAGGTGCCCGGCGCCAGGCTTCATCAGTCCCTGCTTC 2168
524 yArgValMetArgMetLeuAlaValIle.....LeuLeuValV 537
```

```
2169 ACAGGTGGCCATCTGCTGGCACCTTATCTCGGCGCCAGCTGCTCATCGTGG 2218
537 alPheTrpPheLeuIle.....GlyTrpThrSerSerValCysGln 550
2219 TGCCTGGCTGTGTGGAGGACCGGGCACAGGCAAGGAGACAGCCGCC 2268
551 AsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAsp.. 566
2269 GAACGGCGGAGTGTGTGACACTGCGCTGCAACACCGCGATGCAAGTAT 2318
567 .....HisLeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrM 581
2319 GTGGGCTCGCTGGCTACAAATGTCTCTCATC..... 2352
581 etThrAlaValAlaGluPheLeuLeuTrpGlyValTyrLeuCys 597
2353 .....GCCCTCTGCACGCTTTAT...GCC 2373
598 TyrAlaValArgThrValProSerAlaPheHisGluProArgTyrMetAl 614
2374 TTCAATACTCGCAAGTGTCCCGGAAACTTCAACGAGGCCAAGTTCATTGG 2423
614 aValAlaValHisAsnGluLeuIleIleSerAlaIlePheHisThrIleA 631
2424 CTCACCATGTACACCATCATCTGGCTGGCATTTGTGCCCATCT 2473
631 rgPheValLeuAlaSer.....ArgLeuGlnSerAspTrpMetLeuMet 645
2474 TCTATGTACCTCCAGTGTACTACGGGTACAGACCCACCATGTGCGTG 2523
646 LeuTyrPheAlaHisThrLeuThrValThrValThrIleGlyLeuLe 662
2524 .....TCAGTCAGCTCAGCGGCTCCGTGGTGGTGGTGGCTGCCT 2561
662 uLeuIleProLysPhe..... 667
2562 CTTTGGCGCCAAAGTGCACATCATCTCTTCCAGCGCGCAAGAAGCGTGG 2611
668 ..SerHisSerSerAsnAsnProArgAspAspIleAlaThrGluAlaTyr 683
2612 TTAGCCACCGGCGCACCCAGCCCGCTTTGGCAGTGTCTGCCAGGGCC 2661
684 GluAspGluLeuAspMetGlyArgSerGlySerTyrLeuAsnSerSerI1 700
2662 AGCTCCAGCTTGGCCAAAGGTCTGGCTCCCATGTTGTCCCCACTGTGTTG 2711
700 eAsn 701
2712 CAAT 2715
seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ10803
seq_documentation_block:
ID AAZ10803 standard; DNA; 3918 BP.
XX
XX AAZ10803;
XX AC
XX XX
DT 09-NOV-1999 (first entry)
XX
DE Rat metabotropic glutamate receptor 5 (mGluR5) gene.
XX
XX Metabotropic glutamate receptor; mGluR5; central nervous system; CNS;
XX depressant; anaesthesia; phospholipase C; intracellular calcium; ss.
XX
XX Rattus sp.
XX
XX Key Location/Qualifiers
XX CDS 268..3783
XX FT /*tag= a
XX FT /product= mGluR5
XX FT /note= "Metabotropic glutamate receptor 5"
XX
```



```

559 .....IleGlyGlnGlyLysThrSerAspHisLeuI 569
2400 TATCATGGAGCCTCCGGATTAATGCTACTATCAAGCATCCGAGAAG 2449
569 IePheAsnMetCysLeuIleAspArgTrrpAspTyrMetThrAlaValAla 585
2450 TCTACTTGATTGTACACACCACCACTAGGGTGTGTACTCTCTTGA 2499
586 ...GluPheLeuPheLeuLeuTrpGlyValTyrLeuCysTyrAlaValAr 601
2500 TACAATGGATTATTGATTGTTGAGTTCACGTTCTATGCGTTTAAAGACCAG 2549
601 gThrValProSerAlaPheHisGluProArgTyrMetAlaValAlaValIn 618
2550 AAATGTTCCAGCCAACTTAAATGAGCCAAATATATGTTTCCACCATGT 2599
618 isAsnGluLeuIleLeuSerAlaIlePheHisThrIleArgPheValLeu 634
2600 ACACACCTGCATCATATGGCTGGCCCTCGTGCCTATCTACTTT..... 2643
635 AlaSerArgLeuGlnSerAspTrrpMetLeuMetLeuTyrPheAlaHisTh 651
2644 .....GGCAGCACTACAAATATCATCACCATGTTTTCAGT 2681
651 rHisLeuThrValThrValThrIleGlyLeuLeuLeuIleProLys.... 666
2682 CAGCCTCAGTGCACAGTGGCCCTGGGTTGTCATGTTGTCCCGAAGGTGT 2731
667 .....PheSerHisSerSerAsnProArgAspAspIleAlaThr 680
2732 ACATCATCTAGCCAAACCGGAGGAAATGTGGCAGCGCCTTCACAACC 2781
681 Glu...AlaTyrGluAspGluLeuAspMetGlyArgSerGlySerTyrIle 696
2782 TCTACAGTGTACGCATGTCATGTAGGAGACGCAAAATCATCTGCGCTGC 2831
696 u.....AnSerSerIleAsnSerAlaTrpSerGluHisSerLeuAsp 711
2832 CAGCAGATCCAGCAGCTAGTCAACCTGTGGAAGAGAGGGGCTCTCTCTG 2881
711 roGluAspIleArgAspGluLeuLysLys..... 720
2882 GGAACCCCTAGCTCCACGAGGAAATCTGTACTTGGGCCCCAGCAATGAG 2931
721 .....LeuTyrAlaGlnLeuGluIleTyrLysAr 730
2932 AAGAGTACCCGGGGCAACATTTGTGGCAGCAGTCTGTCTCCACATCAA 2981
730 gLysLysMetIleThrAsnAsnProHisLeuGlnLys.....LysA 744
2982 CAAGAAGGAGAACCCCAACCAACGCGGTGCATCAAAACCTTTTCCCAAGA 3031
744 rGysSerLysLysGlyLeuGlyArgSerIleMetArgArgIleThrGlu 760
3032 GCACAGAGAACCCGGGG..... 3048
761 IleProGluThrValSerArgGlnCysSerLysGluAspLysGluGlyAl 777
3049 ...CCTGTGCGACGCGCAGGTGGTGGCTCGGGTCCCGGTGTAGCTGGTGC 3095
777 asPheHisGlyThrAlaLysGlyThrAlaLeuIleArgLysAsnProProG 794
3096 T.....GGTAACGAGGATGCACACACACAGGTGGCCCGGAGCCACCG 3139
794 IuSerSerGlyAsnThrGlyLysSerLysGluGluThrLeuLysAsnArg 810
3140 AGCCGGTCCCAAGGG..... 3156
811 ValPheSerLeuLysLysSerHisSerThrTyrAspHisValArgAspGl 827
3157 .....CTTTATGATGT 3167
827 nThrGluGluSerSerSerLeuProThrGluSerGlnGluGluThrT 844
3168 GCGGAGGAGGAGGAGGAGCTCCCGCGGCTGCCAGCGCGCTCGCCAT 3217
844 hrGluAsnSerThrLeuLeuSerLeuSerGlyLysLysLysLeuThrGlnLys 860
3218 CGCCCATCAGCAGGTGAGCCACCTCGCAGGC..... 3249
861 LeuLysGluAspSerGluAlaGluSerThrGluSerValProLeuValCy 877
3250 .....TCGCGGGCGGCACAGATGATGATGCGCGCTCGTGCA 3287
877 slysserAlaSerAlaHisAsnLeuSerSerGluLysLysThrGlyHisP 894
3288 CTCTGACAGAGCTGCTCGCAGCAGCTGTCGCCAGGCTCGCTCATGGAGC 3337
894 roArgThrSerMetLeuGlnLysSerLeuSerValIleAlaSerAlaLys 910
3338 AGATCAGCAGCGTAGTGCGGCTTCACAGCCAAATCTCCGAGCTCAAC 3387
911 GluLysThrLeuGlyLeuAlaGly..... 918
3388 TCCATGATGTTGTCACCGCGGCCAGCGCGGCGGCGGCTGGTACTCCAAT 3437
919 .....LysThrGlnT 922
3438 CTGCTCCTCTACTCTATCCCAAGAGATTGAGCTGCCAGCAGCATGA 3487
922 hrAlaGlyValGluGluArgThrLysSerGlnLysProLeuProLysAsp 938
3488 CGACCTTCGACAG.....ATCCAGCGCTGCGCGGCATC 3522
939 LysGluThrAsnArgAsnHisSerAsnSerAspAsnThrGluThrLysAs 955
3523 -GAGGTGACC.....GGAGAGCTCAGGGGGCGACAGGCGTATC 3560
955 proAlaProGlnAsnSerAsnProAlaGlu...ProArgLysProG 971
3561 ACCTGCCAGGAGCGCCACAGAGAGCTGAATCGCGCCCGGCAACACAG 3610
971 InLysSerGlyIleMetLysGlnArgValAsnProThrThrAlaAsn 987
3611 ATCTGGAGGAGCTTGGCGCTCACTCCACCATCGCCC..... 3648
988 SerAspLeuAsnProGlyThrThrGlnMetLysAspAsnPheAspIleGl 1004
3649 .....TTCAGGAGCTCGTGGACTCGG 3671
1004 yGluValCysProTrrpGluValTyrAspLeuThrProGlyProValProS 1021
3672 GAGCACCACCCCAACTCTCCAGTCTCAGAAATCGGCCCTCTCATCCCAT 3721
1021 erGluSerLys 1024
3722 CCTCTCCCAAA 3732
seq_name: /SIBS1/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ80420
seq_documentation_block:
ID: AAQ80420 standard; cDNA; 4181 BP.
XX
AC AAQ80420;
XX
DT 21-JUL-1995 (first entry)
XX
DE Human mGluR5b cDNA.
XX
KW Metabotropic glutamate receptor; mGluR5; mGluR5b; hippocampus;
XX cerebellum; pCMV-T7-3; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 370..4008

```


782 AlaLysGlyThrAlaLeuIleArgLysAsnProGluSerSerGlyAs 798
 : :: ||| ||| :::: |::| |::|||
 3304 gTGGCGGTCCAGGCCGCGCCGAGGGGGCGAGTCCTCCACAGC 3353
 798 nThrGlyLysSerLysGluGluThrLeuLysAsnArgValPheSerLeuL 815
 : ::||| :::
 3354 CGGCCCAAGAGCG 3366
 815 ySLysHisSerThrTyrAspHisValArgAspGlnThrGluGluSer 831
 : :: ||| ::||| :::
 3367CTGTATGATGGCCGAGCGTGAG 3390
 832 SerSerLeuProThrGluSerGlnGluGluThrThrGluAsnSerTh 848.
 : ::||| ::::: ::::: :::
 3391 GAGCACTTCGGCGCGCGCGCGCGCTCACCGTCGCCCATCAGCAC 3440
 848 rIeuGluSerIeuSerGlyLysLysLeuThrGlnLysLeuLysGluAsps 865
 : ::||| ::::: :::
 3441 GCTGAGCCACC CGCGGCTCGGCCAGCGCCAGCGAGC 3478
 865 eRGLuaGluSerThrGluSerValProLeuValCysLysSerAlaSer 881
 : ::||| ::||| |::||| ::::: |::|||
 3479 ACGATGTCCGTCGCTGCACNCGAGCGCTGTGGCGGCACACTCTCTCG 3528
 882 AlaHisAsnLeuSerSerGluLysLysThrGlyHisProArgThrSerMe 898
 : ::||| :::
 3529CAGGCGTCCCT 3539
 898 tIeuGlnLysSerLeuSerValIleAlaSerAlaLysGluLysThrLeuG 915
 : ::||| ::||| ::||| :::
 3540 CATGAGCAGATCAGCATGTGGTCACCCGCTTC 3573
 915 LyLeuAlaGlyLysThrGlnThrAlaGlyValGluGluArgThrLysSer 931
 : ::||| ::||| ::||| ::|||
 3574ACGCCAACATCAGCCAG 3591
 932 GlnLysProLeuProLysAspLysGluThrAsnArgAsnHisSerAsnSe 948
 : ::||| ::||| ::||| ::|||
 3592CTCACTC 3599
 948 rAspAsnThrGluThrLysAspProAlarProGlnAsnSerAsnProAla 964
 : ::||| ::||| ::||| ::|||
 3600 CATGATGTCTCCACCGCGCCCGCCAGCCCGCGTGGCGGCCGCTCT 3649
 965 GluGluProArgLysProGlnLysSerGly..... 974
 3650 GCTGTCTCTACTGATCCCCAAGAAGATCCAGTTGCCCGCAGCATTGAGC 3699
 975IleMetLysGlnGlnArgValAsnProThrThrAlaAsnSerAspL 990
 : ::||| ::||| ::||| ::|||
 3700 ACCITTCGCGAAATCCAGCTCTCCCGGCCATCGAAGTCACGGCGGCC 3749
 990 euAsnPro.....GlyThrThrGlnMetLysAsp 999
 : ::||| ::||| ::|||
 3750 TCAGCCCGCGCAGGGCGAGCGGTGGGGACGCGCGCGCGGAGAGCC 3799
 1000 AsnPheAspIleGlyGluValCysPro.....TrpGluValTy 1012
 3800 CGCGCGCGGTCCCGAGGCTGCGGCCGCCCAAGCCAGCACTGGAGGAGCTG 3849
 1012 rAspLeuThrProGlyProValProSerGluSerLysValGlnLysHisV 1029
 : ::||| ::||| ::||| ::|||
 3850 GTGGCTCTCACCCCGCGTCCCTTCAGAGACTCGGTGGACTCGGGGAG 3899
 1029 alSerIleValAlaSerGluMetGluLysAsnProThrPheSerLeuLys 1045
 : ::||| ::||| ::||| ::|||
 3900 CACAACCCCAACTCGCAGTGTCCGAGTCGGCGCTCTGTATCCCGCTGT 3949
 1046GluLysSerHisLysProLy 1053
 : ::||| ::||| ::|||
 3950 CTCCCAATAATGACACTCTTATCATTAAGAGATTACACTCAGAGCTCTCTCG 3999
 1053 sAlaAlaGluValCysGlnGlnSerAsnGlnLysArgIleAspLysAlag 1070

```

i      : ::::|
4000 TCGTGTGAA..... 4009
1070 luValCysLeuTrpGluSerGlnGlnSerIleLeuGluAspGluLys 1086
    ||| ::::|
4010 .....TGTCCCTGAAAGCAGC..... 4027
1087 LeuLeuIleSerLysThrProValLeuProGluArgAlaLysGluGluAs 1103
    ||| ::::|
4028 .....CGCCCTGCGGTGCG..... 4042
1103 nGlyGlnProArgAlaAsnValCysAlaGlyGlnSerGluGluL 1120
    ||| ::::|
4043 .....GAGCGGA 4049
1120 euProProLysAlaValAlaSerLysThrGluAsnGluAsnLeuAsnGln 1136
    ||| ::::|
4050 GCCCCCG.....TGTTCCACACACACAT..... 4075
1137 IleGlyHisGlnGluLysLysThrSerSerSerGluGluAsnValArgG1 1153
    ||| ::::|
4076 ...GGCAAGCATAGTCGCTGGTTACGCGCCAGGGGGAAGATGCCAAGG 4122
1153 ySer 1154
    |||
4123 CACC 4126

```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV21511

seq_documentation_block:

ID AAV21511 standard; DNA; 6755 BP.

XX AC AAV21511;

XX DT 17-AUG-1998 (first entry)

XX DE Staphylococcal bacteriocin BacRI operon.

XX KW BacRI; bacteriocin; antimicrobial; antibacterial; antibiotic;
 KW Moraxella bovis; infectious bovine keratoconjunctivitis; cancer;
 KW therapy; ds.

XX OS Staphylococcus aureus strain UT0007 (ATCC 55800).

XX PN W09812319-Al.

XX PD 26-MAR-1998.

XX PF 18-SEP-1997; 97WO-US16758.

XX PR 17-SEP-1997; 97US-0931999.

XX PR 19-SEP-1996; 96US-0710561.

XX PA (UNIV) UNIV KANSAS STATE RES FOUND.

XX PI Crupper SS, Iandolo JJ;

XX DR WPI; 1998-230316/20.

XX PT Therapeutic proteinaceous substances from *Staphylococcus aureus* -
 PT useful to inhibit growth of wide range of prokaryotic or eukaryotic
 PT cells, e.g. *Moraxella bovis* causing infectious bovine
 PT keratoconjunctivitis

XX PS Claim 2; Page 19-23; 38pp; English.

XX CC This polynucleotide comprises the bacteriocin BacRI operon of
 CC *Staphylococcus aureus* UT0007. The sequence of the BacRI operon
 CC was determined by N-terminal sequencing of purified BacRI peptide
 CC (see AAW54171), with back-translation and plasmid analysis. The
 CC BacRI operon includes the BacRI gene (see AAV21510), a homologue of
 CC the *cylM* gene of the cytolysin operon of *Enterococcus faecalis*,
 CC whose function is involved in the maturation of pre-cytolysin,

CC an ATP-transporter gene, *bio1* and *bio2* genes related to
 CC lactococcal biosynthesis and modification, and a gene involved in
 CC immunity function. BacRI peptide can be produced by construction
 CC of an expression vector containing an oligonucleotide or operon
 CC coding for BacRI, and use of the vector to transform host cells for
 CC BacRI expression. The entire BacRI operon has been cloned into
 CC plasmid PUB110, and *Bacillus subtilis* transformants secreted the
 CC recombinant BacRI peptide into the medium. Bacteriocin BacRI is
 CC active against many Gram-positive and Gram-negative organisms such
 CC as *Bordetella bronchiseptica*, *Pasteurella multocida* and
 CC *Staphylococcus aureus*; *Moraxella bovis*, causing infectious bovine
 CC keratoconjunctivitis, is especially sensitive. BacRI can also be
 CC used as an anti-cancer agent.

XX SQ Sequence 6755 BP; 3903 A; 1252 C; 1600 G; 0 U; 0 other;

alignment_scores:

Quality: 204.50 Length: 461
 Ratio: 0.867 Gaps: 20
 Percent Similarity: 51.193 Percent Identity: 22.993

alignment_block:

US-09-775-181-2 x AAV21511

Align seg 1/1 to: AAV21511 from: 1 to: 6755

```

717 GluLeuLysLysLeuTyAlaGlnLeuGluIleTyLysArgLysLysMe 733
    ||| ||||| ||||| ::::|
3609 GAAAAAAMAAACCCCAAGCACACACCGACGACCAAAAAA..... 3653
733 tIleThrAsnAsnProHisLeuGlnLysLysArgCysSerLysLysGlyL 750
    ||| ::::|
3654 ....ACGCCGAAACCAACAGAGAAACCAAAAGACAGAGAGAAAAA 3699
750 euGlyArgSerIleMetArgIleThrGluIleProGluThrValSer 766
    ||||| ||||| |||||
3700 AAAACCGGAGCAAAACGCGAGAAACGAG.....GAACAAAGAAAA 3743
767 Arg.GlnCysSerLysLysGluAspLysGluTyAlaAspHisGlyThrAlaL 783
    ::::| ||| ::::|
3744 AACCACAAAAAGGAAAAAGCAAAACCAACCGGACAGACCAACAAAAA 3793
783 ysglyThrAlaLeuIleArgLysAsnProProGluSerSerGlyAsnThr 799
    ||||| ::::|
3794 AAGGCACAGAGAGAAAGCAAAAGAAAAACCAACAAAGAGAACGACGAA 3843
800 GlyLysSerLysGluGluThrLeuLysAsnArgValPheSerLeuLysLy 816
    ||:::|
3844 AAGAAAAACAGGAAGAGCGCAACAAACCAAAACCGCACAAAGAGAGAG 3893
816 sSerHisSerThrTyAspHisValArgAsp...GlnThrGluGluSerS 832
    ::::| ::::|
3894 AAGGAGCAGCAAGCAACCAACCAAAAGAGACCCCAAGCAACAAAGGCA 3943
832 erSerLeuProThrGluSerGlnGluGluThrThrGluAsnSerThr 848
    ||:::| ||| ::::|
3944 CGCAACGCCAGGCGCACAGACAGAGAGAGAAAAACAAACCAACCAACG 3993
849 LeuGluSerLeuSerGlyLysLysLeuThrGlnLysLeuLysGluAspSe 865
    ::::| ::::|
3994 AGGAACAGCGCGCGGAAAAAAGAAAAAAGAGAGAGAGAGAAAAA 4043
865 rGluAlaGluSer..ThrGluSerValProLeuValCysLysSerAlaSer 881
    |||||
4044 CCAAGCAGACACAAACCAAAAA..... 4066
882 AlaHisAsnLeuSerSerGluLysLysThrGlyHisProArgThrSerMe 898
    ||| ::::|
4067 ...AAGAACAGCAGAGAGAGGCAAAACCCACCCCAACCAACAAAAA 4113
898 tLeuGln.LysSerSerValIleAlaSerAlaLysGluLysThrLeu 914
    ||| ||:::|

```


320 TCATGGCGCTCATGCCGCTCAACAGGAGGTGGCCAAAGGGCAGCATCGGG 369

49ProGlyArgAlaSerAlaSerA 56
||||: ||||||:|

370 CGCGCGCTCTCCCGCGCTGGAGCTACCCATCAGCAGATCCGCAACGA 419

56 sPserSer..AlaPro..... 60
:: ||| |||||

420 GTCACCTCTGGCGCCCTACTTCTTGGACCTCGCACTCTATGACACCGAGT 469

61TrpS 62
||||: |||

470 GTGACATGCNAAGGAGCTGAAGCCCTTCTATGACGCAATAAGATGGG 519

62 erArgSerThrAspGlyThrIleLeuAlaGlnLysLeuAlaGluGluVal 78
::: |||||:|

520 CGCAACCATTTGATGCTTTGGAGGCGTCTCCGCTCTGTCACATCTAT 569

79 ProMetAspValAlaSerTyrlLeuTyrrThrGlyAspSerHisGlnLeuLy 95
||| ||| :||| :|

570 TATCGCGGAGTCCCTCCAAGGCTGAACTCTGTCGACGTTTCCTTCGCG 619

95 sArgAlaAsnCysSer..... 100
::: |||||

620 CCACCAACGCGCTCTCTCGGATAGAAGAAGTACCCGATATTCTTCGCG 669

101 ..GlyArgTyrlGluLeuAlaGlyLeuProGlyLysTrpProAlaLeuAla 116
||| :| :| ||| |||

670 ACGGTCCGCTCAGCAACCCGTTGAACCC..... 698

117 SerAlaHisProSerLeuHisArgAlaLeuAspThrLeu..... 129
|||||:|

699 ...CGCCATCTCTGAAGCTCTCTGAAGCACTTCCGTCGGCGCTGGCGCA 745

130 ThrHisAla..... 132
||| |||||

746 CACTCAGCAGGAGCGYGCAGCGCTTCTCCGAGGTGAGGAATGACCTGACT 795

133ThrAspPheLeu 136
:| :| :| :| :| :| :|

796 GGGTTCTCTATGGGAAGATATTGAGATCTCAGACACACAGAGATTCTC 845

137 AsnValMetLeuGlnSerAsnLysSerArgGluGlnAsnLeuGlnAspAs 153
::| :| :| :| :| :| :| :| :| :|

846 CAATGATCCTGCACAGCGTCAAAAAGCTCAAGGGGAATCAGCTGGCGGA 895

153 pIeuAspTrp..TyrlAlaLeuValTrpSerLeuLeuGluGlyGluPr 169
||| ||| :| :| :| :| :| :| :| :| :|

896 TCATCTTGGCGAGTTGACCAAGATATGGCAAAAGTCTTCTGTGTGT 945

169 oSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeuSerAlap 186
||| :| :| :| :| :| :| :| :| :|

945 GCCTTCGAGGAGAGCATGTTGGCAGCAAGTACCAGTGGATCATCCCGGG 995

186 roAlaProGlnValPheLeuGlnAlaThrArgGluGluSerArgIleLeu 202
:: :| :| :| :| :| :| :| :| :|

996 ATGTTACGAGCGTG.....CCTGGTGGGAGCAGGTGCATG 1030

203 LeuGlnAspLeuSerSerSerAlapProHisLeuAlaAsnAlaThrLeuGl 219

1031 TGGAGGCCAATT.....CCTCAGCTCCCTGCACA.....GA 1062

219 uThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHis 235
| :| :| :| :| :| :| :| :| :|

1063 AGCTTCCTGGCTGCCATGGA.....AGGTTACAT 1091

236 ArgArgGly.....ProAsnGlnGlyProArgGlyLeuGlyHis 249
||| ||| :| :| :| :| :| :| :| :|

1092 CGGAGTGGACTTTGAGCCCTGAG..... 1115

249 rTrpArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrp 265
||| :| :| :| :| :| :| :| :| :|

1116CTCCAACAACAAATCAAGACACTCTCAGGAAG 1146

[illegible]

```
525 ArgValMetArgMetLeuAlaValIleLeuLeuValValPheTrpPheLeu 541
1957 .....ATGCTGCTCATGGACCTGTGTCATCCT 1982
541 uileGlyTrpThrSer..... 546
1983 GATCTGTGGCAGGCTGTGGACCCCTGGCGAGGACAGTGGAGAGGTACA 2032
546 ..... 546
2033 GCATGGAGCGGACCCAGCAGCGCGGACATCTCCATCCGCCCATTTGCTG 2082
547 SerValCysGlnAsnLeuGluLysGlnIleSerLeuIleClyGlnGlyLy 563
2083 GAACACTGGGAAACACCCACATGACATC..... 2112
563 sThrSerAspHisLeuIlePheAsnMetCysLeuIleAspArgTrpAspT 580
2113 .....TGGCTTG 2119
580 yrMetThrAlaValAlaGluPheLeuPheLeuLeuTrpGlyValTyrLeu 596
2120 GCATTGCTACGCTACAGGGCTCCTCATGCTATTGCTGTTCTTCTG 2169
597 CysTyrAlaValArgThrValPro....SerAlaPheHisGluProArgTyr 612
2170 GCATGGGAAACCCGCAATGTGACATCTGCCCTCAAGCAGACAGCAAGTA 2219
612 rMetAlaValAlaValHisAsnGluLeuIleIleSerAlaIlePheHisT 629
2220 CATCGGCATGAGTGTACAATGTGGGGATCATGTCCATCATCTCGGGGCTG 2269
629 hrIleArgPheValLeuAlaSerArgLeuGlnSerAspTrp..... 642
2270 CTGTCTCTCTTCGACGCGTGACACGCCCAACGTCGAGTCTGTCATCGTG 2319
643 MetLeuMetLeuTyrPheAlaHisThrHisLeuThrValThrValThrI 659
2320 GCCCTGGTCATCTCTTCGACG.....ACCATCACTCT 2354
659 eGlyLeuLeuLeuIleProLysPheSerHisSerSerAsnAsnProArgA 676
2355 CTGCTCTGGTGTTCGCCAAGCTCATCACTCTGAGGACAAACCCCT.... 2400
676 sPAspileAlaThrGlu.....AlaTyrGluAspGluLeuAspMetGly 690
2401 ..GAGCGACGCACTCAGACAGCGGTTCCAGTTCACACAGAACCCAGAG 2448
691 ArgSerGlySerTyrLeuAsnSerSerIleAsnSerAlaTrpSerGluHi 707
2449 AAGAAGATTCGAAGACCTCCACTTCAGTCACACGCGTGAACGACGCGAG 2498
707 sSerLeuAspProGluAspIleArgAspGluLeuLysLysLeuTyrAlaG 724
2499 CAGCTCACCCCTGGAGGACTCGAGTCAGAACCAACCCGCCCT..... 2541
724 InLeuGluIleTyrLysArgLysLysMetIleThrAsnAsnProHisLeu 740
2542 .....CGAATGAAGATCACAGAGCTGGGACAAAGACTTG 2574
741 GlnLysLysArgCysSerLysLysGlyLeuGlyArgSerIleMetArgA 757
2575 GAAGAAGTCACCATCGACTCAAG.....ACACACCCAGAGAA 2612
757 rgIleThrGluIleProGluThrValSerArgGlnCysSerLysGluAsp 773
2613 GA.....CCACATACATCAACAGAAATCACTACCRAGAGCTCA 2650
774 LysGluGlyAlaAspHisClyThrAlaLysGlyThrAlaLeuIleArgLy 790
2651 ACGACATCCTCAGCTGGGCACTTCACAG...AGAGCACACATGGAGGA 2697
790 sAsnPro.....ProGluSerSerGlyA 798
```

```
2698 AGGGCCATTCTAAAAATCACCTCGATCAAAACCCCCAGCTCCAGTGA. 2746
798 snThrGlyLysSerLysGluGluThrLeuLysAsnArgValPheSerLeu 814
2747 ..ACAGCAGAGCCCTCAAGAACATGCAAGACC...CCATAGAAGACA 2791
815 LysLysSerHisSerThrTyrAspHisValArgAspGlnThrGluLysE 831
2792 TCAACTCCCCGGGACACATCCAGCGCGGCTGCTCCAGCTCCCATC 2841
831 rSerSerLeuProThrGluSerGlnGluGluThrThrGluAsnSerT 848
2842 CTTCAACGCGCTAGCTCCATCCATCGGAGGCTGGAGTCCAGCTGCG. 2890
848 hrLeuGluSerLeuSerGlyLysLysLeuThrGlnLysLeuLysGluAsp 864
2891 .....TCAGCCCTGTGTCAGCCCTACCGCCAGCCCTGCCACAGAC 2932
865 SerGluAlaGluSerThrGluSerValProLeuValCysLys..... 878
2933 ACGTACCACCCCTCCTCCGAGTCTCGGGCTGTAGGGGTGGGAG 2982
879 .....SerAlaSerAlaHisAsnLeuSer..... 886
2983 GCCTGGGCGGGGCTCCCGGGACAGCACCATGCTGGGCCAAGCGCC 3032
887 .....SerGluLysLysThrGlyHis..... 893
3033 TGCCACAGGCACACTGACGGCGGAGAGTGGGCACCATCTGCCTCT 3082
894 .....ProArgTh 896
3083 CCAGACTGCTGAATGGCGCTCAGCGAGAGCGGACTCGGCACCGACCTC 3132
896 rSerMetLeuGlnLysSerLeu..... 903
3133 GAGCCTTATCTGTGAAGTCTTACTCTCACAGAGGAGGAATGACAATG 3182
904 .....SerValIleAlaSerLysGluLysThrLeuGly..... 915
3183 ACTTCTCTTCTTGGCGTCTGCAAAACAAAGAGGAGTGGGATGCTGAAA 3232
916 LeuAlaGlyLysThrGlnThrAlaGlyValGluArgThrLysSerGl 932
3233 CTGCAANAAACAAATCAACCTAGACAAAGAGAGAGCGCTCGGACTCC 3282
932 nLysProLeuProLysAspLysGluThrAsn.....ArgAsn...HisS 946
3283 TGCTCTCTCGCCAAAGTGGCCAGAGCAAGGCTCTGCAGAAATGCCCTC 3332
946 erAsnSerAspAsnThrGluThrLysAspProAlaProGluAsnSerAsn 962
3333 GGCACCTCTGCACACACGTCGCGTGGTGGCGGCCGCCCCACACAC..... 3376
963 ProAlaGluGluProArgLysProGlnLysSerGlyIleMetLysGlnGl 979
3377 .....CAGCCCTTCTCCACTCTGCGGGGAGAGGTGAG 3408
979 nArgValAsnProThrThrAlaAsnSerAspLeuAsn.....Prog 993
3409 CAGGCGCCACAGGTCCCGCAGAGCTGGGCCAGTCCCATCTTCAGACTTCCTG 3458
993 lYThrThrGlnMetLysAspAspPheAspIleGlyGluValCysPro 1008
3459 GAACAAGAGGAGTACAGCAGAAAG.....GTAGCCGGCTCTGCGCT 3499
```

seq_name: /SIDS1/gcdata/geneseq/geneseqn/NA2000.DAT:AAZ50616
seq_documentation_block:
ID AAZ50616 standard; cDNA; 2922 BP.
XX
AC AAZ50616;

```

XX 20-JUN-2000 (first entry)
XX Human calcium sensing receptor isoform, CaSRd encoding cDNA.
XX
XX Calcium Sensing Receptor; CaSRd; isoform; human; splice variant;
XX hypotensive; osteopathic; receptor activity; calcium level; modulator;
XX treatment; hyperparathyroidism; osteoporosis; Paget's disease;
XX hypercalcaemia malignancy; hypertension; gene therapy; ss.
XX
XX Homo sapiens.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 1..2922
XX /tag= a
XX /product= "Human calcium sensing receptor isoform, CaSRd"
XX /note= "Does not include stop codon"
XX /partial
XX
XX W0200006601-A1.
XX
XX 10-FEB-2000.
XX
XX 28-JUL-1999; 99WO-US17116.
XX
XX 30-JUL-1998; 98US-0094702.
XX
XX (AVET ) AVENTIS PHARM PROD INC.
XX
XX Yu KT, Labaudiniere RF, Thrower LW;
XX
XX WPI; 2000-195263/17.
XX
XX P-PSDB; AAY45000.
XX
XX Nucleic acids encoding isoforms of human calcium sensing receptor for
XX treating, e.g. hyperparathyroidism or osteoporosis
XX
XX Claim 8; Page 60-66; 81pp; English.
XX
XX The present sequence is the cDNA encoding an isoform of the human
XX calcium sensing receptor CaSRd, expressed in the kidney. It is a splice
XX variant of the wild type CaSRd, that arise from deletion of nucleotides
XX 1080-1391, from the extracellular domain, that comprise acidic residues.
XX It has hypotensive and osteopathic activity. The CaSR isoforms can be
XX used to identify agonists and antagonists that modulate the receptor
XX activity and calcium levels. These modulators are useful for treating
XX hyperparathyroidism or osteoporosis. Paget's disease, hypercalcaemia
XX malignancy or hypertension. The DNA sequence is also useful for altering
XX the CaSR activity and in gene therapy.
XX Note: This sequence has deletion of nucleotides 1080-1391, according to
XX the sequence shown in the specification, but has been stated as
XX 1075-1386 in the claims.
XX
XX Sequence 2922 BP; 555 A; 848 C; 752 G; 667 T; 0 other;

```

```

alignment_scores:
  Quality: 200.50      Length: 941
  Ratio: 0.466        Gaps: 40
  Percent Similarity: 45.696  Percent Identity: 19.766

```

```
alignment_block:
```

```
US-09-775-181-2 x AAZ50616
```

```
Align seg 1/1 to: AAZ50616 from: 1 to: 2922
```

```

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
   ::::::::::::::::::::
394 TCAGAGCACATTCCCTCTACGATTGCTGGTGGAGCAACTGGCTCAGG 443
   ::::::::::::::::::::
91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
   :   :   :   :   :   :   :   :   :   :   :   :   :

```

```

444 CGTCTCCACGGCAGTGGCAAAATCTGCTGGGGCTCTTCTACATT..... 486
108 euProGlyLysTyrProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
   ||| ::::: ||| ::::: |||
487 .....CCCCAGGTCAGTTATGCTCTCCACG.....AGA 516
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG1 141
   ||| ::::: ::::: ::::: ::::: ::::: :::::
517 CTCCTCAACAACAAGAATCAATTCAGTCTTCTCTCCGAACC...ATCCC 563
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAspTyrG 158
   ::::: ::::: ::::: ||| ::::: :::::
564 CAATGATCAGCAGCAGCCAGCTGCCATGCCAGACATCATCGAGTATTTCC 613
158 InAlaLeuValTyrSerLeuGluGluProSerIleSer..... 172
   : : : : : ||| ::::: :::::
514 GCTGGAACCTGGTGGGCACAATTGCACTGATGACGACTATGGGGCGCG 663
173 .....ArgAlaAlaIleThrPheSerT 180
   ||| ::::: |||
564 GGGATTGAGAAATCCGAGAGGAAGCTGAGGAAGGGATATCTGCATCGA 713
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
   ||| ::::: ||| :::::
714 CTTCTGCACTCATCTCCAGTACTCTGATGAGGAAGAGATCCAGCATG 763
196 .....ArgGluGluSer.ArgIleLeuLeuG 204
   ||| ::::: ||| :::::
764 TGGTAGAGGTGATTCAAAATCCAGCGCCAAAGTCATCGTGTTCCTCC 813
204 InAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
   ||| ::::: ||| :::::
814 AGTGGCCAGATCTTGAGCCCTCATCAAGAGAGATTGCGCGCGCAA... 860
221 GluTyrPheHisGlyLeuArgArgLysTyrPArgProHisLeuHisArgAr 237
   ::::: ||| ::::: |||
861 .....TATCAGCGGCAAGATCTGGCTGGCAGCGAGCGCTGGCGCCAGCT 904
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
   ||| ::::: |||
905 CCTCCTGATGCCATGCTCAGTACTTCCAGTGGTGGCGGCACCAT 954
250 rpArgArgLysAspGlyLeuGlyAspLysSerHisPheLysTyrPser 266
   ||| ::::: |||
955 GGATTGCTCTGAAGC.....TGGGCA 977
267 ProProTyrLeuGlu.....Cy 272
   ||| ::::: |||
978 GATCCAGCGCTTCGGGGAATTCCTGAGAAGGTCCATCCAGGAAGTCTG 1027
272 s.GluAsnGlySerTyrLysProGlyTyrLeuValThrLeuSerSerAla 288
   : : : : : ||| :::::
1028 TCCACATGGTTTCCCAAGGAGTTTGGGAAGAACAATTTAACTGCCAC 1077
289 IleTyrGlyLeuGlnProAsnLeuValProGluPheArgGlyValMetLy 305
   : : : : : |||
1078 CTACGGCATCTA.....AACTTACAAACAATATGGG 1109
305 sValAspIleAsnLeuGlnLysValAspIleAspGlnCysSerSerAspG 322
   ::::: ||| ::::: |||
1110 G.....GAGCAGGTGACCTTTGATGAGTGT.....G 1135
322 LyrTrpPheSerGlyThrHisLysCys.....HisLeuAsnAsnSer 335
   ||| ::::: |||
1136 GTGACCTGGTGGGNACTATTCCATCATCACTGGCACCTCTCCCGAG 1185
   ||| ::::: |||
1186 GATGGCTCCATCGTGTTAAGGAAGTCGGGTAT.....TACAA 1223
351 uCysIleCysLysAlaGly.....PheTyrHisProGlyValLeuP 365
   : : : : : |||
1224 CGTCTATCCCAAGAGGGAAGAAAGACTCTTCTCAACGAGGAGAAATCC 1273

```

[illegible]

||||:| :|||
2882 GTGGTGGAGGCGACT 2898

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AA86167

seq_documentation_block:

ID AA86167 standard; DNA; 3219 BP.

XX AA86167;

XX 03-DEC-1997 (first entry)

XX Nucleotide sequence of pratch3.

XX Chimeric receptor; extracellular domain; seven transmembrane domain;
KW intracellular cytoplasmic tail domain; metabotropic glutamate receptor;
KW MGR; calcium receptor; CR; mGluR; G protein-coupled receptor;
KW neurological disease; ss.

XX Chimeric - Rattus rattus.

XX Key Location/Qualifiers
FH 25..3195
FT CDS
FT /*tag= a
FT /product= pratch3
XX

XX M090705252-A2.

XX 13-FEB-1997.

XX 25-JUL-1996; 96WO-US12336.

XX 26-JUL-1995; 95US-0001526.

XX (NPSP-) NPS PHARM INC.

XX Fuller FH, Hammerland LG, Krapcho KJ;

XX WPI; 1997-145690/13.

XX P-PSDB; AAW25764.

XX Chimeric receptors comprising metabotropic glutamate receptor and
PT calcium receptor - used for screening for neurologically active
PT compounds

XX Example 4; Fig 4; 177pp; English.

XX This sequence encodes a chimeric receptor. The chimeric receptor
CC comprises an extracellular domain, a seven transmembrane domain, and
CC an intracellular cytoplasmic tail domain, and a sequence of at least
CC 6 contiguous amino acids is homologous to a sequence of a metabotropic
CC glutamate receptor (MGR), and a sequence of at least 6 contiguous amino
CC acids is homologous to a sequence of a calcium receptor (CR). The
CC chimeric receptor may be used for screening for compounds that bind to
CC or modulate the activity of MGR or CR and for determining the site-of-
CC action of a CR active compound. The compounds can be used in the
CC treatment of neurological diseases and disorders. They can also be
CC used as diagnostic agents. Chimeric receptors such as this, allow the
CC coupling of certain functional aspects of an MGR with certain functional
CC aspects of a CR. They allow for more efficient high-throughput screening
CC of compounds.

XX Sequence 3219 BP; 775 A; 880 C; 851 G; 713 T; 0 other;

alignment_scores:

Quality: 199.50 Length: 964
Ratio: 0.503 Gaps: 43
Percent Similarity: 41.183 Percent Identity: 20.228

alignment_block:

US-09-775-181-2 x AA86167 ..

Align seg 1/1 to: AA86167 from: 1 to: 3219

```

153 AspLeuAspTrpTyrGlnAlaLeuValTrpSerLeuLeuGluGlyGluPr 169
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
501 GATCGGCCCTGGCTCCAGCTGTGGCCATTCAGTCCAGAATCTTCTCC 550
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169 oSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeu..... 183
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
551 A.....GCTGTTCGACATCCACAGATCGCTATTCCTGCC 585
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 .....SerAlaProGlnValPheLeuGlnAlaThrArgGlu... 197
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
586 ACAAGCATAGACCTGAGTGACAAACTTTGACAAATACCTTCCTGAGGT 635
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
198 .....GTTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 201
636 GGTCCCTCTGACACTTTGACAGGCAAGGCGATGCTCGACATACCTCAAGC 685
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 eLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrL 218
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
686 GTTACAACTGGACCTATGCTCAGCAGTCCACAGAGGAGGAA..... 728
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 euGluThrGluTrpPheHisGlyLeuArgArgLysTrp..... 230
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
729 .....TTACGGCAGAGTGGAAATGGATGGATGCTTT 755
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
231 .....ArgProHisLeuHisArgArg...GlyPr 239
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
756 CAAAGAACTGGCTGCCCGAGGAGGCT...CTGCATCGCACACTCGGACA 802
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
239 OAsn..... 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
803 AAATCTACAGCAATGCTGGCGAGAGAGCTTTGACGGGCTCCTCGCTAAA 852
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 .....GloglyProArgGlyLeuGlyHisSerTrpArg 251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
853 CTCGGGAGGGCTTCCCAAGGC...CAGGGTGTGTGCTCTCTCTCGCA 899
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
252 ArgLysAspGlyLeuGlyAspLysSerHis..PheLysTrp..... 265
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
900 GGGCATCACAGTGGGGCTTACTGAGTGCATGGCGCGCTGGCGCTCG 949
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 .....SerProTyrLeuGluCysGluAsnGlySer..... 276
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
950 TGGCGAGTTCCTCATTGGAAGTGTGATGGATGGCGAGACAGAGATGAA 999
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
277 .....TyrLysPro..... 279
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1000 GTCATCGAAGGCTATGAGTGGAAAGCCACGGAGGATCAATAAGCT 1049
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
280 .....Gly 280
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1050 TCAGTCTCCAGAGTCAAGTCAATTTGATGACTACTTCTTGAAGCTGAGC 1099
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
281 TrpLeuValThr.....LeuSe 286
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1100 TGGACACCAACACAGGAATCCTTGGTTCCCTGAGTTCGGCAACATCGC 1149
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 rSerAlaIleTyrGlyLeuGln.....ProAsnLeu..... 296
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1150 TTCAGTGTGCGCTTACCTGGACCTCTTGGAAACCCCACTTTAAGAA 1199
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
297 .....ValProGluPheArgGlyValMetLysValAspIleAsnLeuGln 311
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1200 AGTGTGCACAGGAATGAAGCTTGGAAAGAAACTATGTCAGGACAGCA 1249
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
322 LysValAspIleAspGlnCysSerSerAspGlyTrpPheSerGlyThrH1 328
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1250 AAATGGGATTTGTCAATCAATGCCATCTATGCCATGGCACATGGCTGCAG 1299
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 sLysCys..His.....LeuAsnAsnSerGluCysMe 338
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1300 AACATGCACCATGCTCTGTGTCCCGGCATGTGGGCGCTGTGTGATGCTAT 1349
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

3033 GACCGTCAGGAACAGTCTGCAAGGACCTGTGGTGGAGACCAGCGG. 3081
      ::||| |||::: ||| ::: ::::
817  erHisSerThrTyAspHisValArgAspGlnThrGluGluSerSerSer 833
      ::|||::: ||| ||| :::
3082 .....CCAGAGTGGAGGACCCCTGAAGAGTTGTCCCCACCA 3117
      ::|||::: ||| ||| :::
834  LeuProThrGluSerGlnGluGlu.....GluTh 843
      ||| ::::|||::: ||| :::
3118 CTTGTAGTGTCCAGTTCACAGAGCTTTGTCATCAGTGTGGAGGCAGCAC 3167
      ||| ::::|||::: ||| :::
843  rThrGluAsnSerThrLeuGluSerLeuSerGlyLysLys 856
      |::: ::::|||::: ||| :::
3168 TGTTACAGAAAACGTAGTGAATTCTATAAATGGAAGGAGA 3207
      |::: ::::|||::: ||| :::

```


91 rHISGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
: : : : :
484 CGTCTCCAGCGGAGTGGCAAACTGCTGGGGCTCTCTACATT 526
108 euProGlyLysrppProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
: : : : :
527CCCGAGGTCAGTTATGCTCTCCAGCAGA 556
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuCl 141
: : : : :
557 CTCTCAGCAACAGNATCAATCAAGTCTTCTCCCGAACCCATCCC 503
141 nSerAsnLysSerArgGluGlnAnLeuGlnAspLeuAspLysrptTyrG 158
: : : : :
604 CAATGATGAGCACCAGCGGCACCTGCCATGGCAGACATCATCGAGTATTCC 653
158 lnAlaLeuValTrpSerLeuLeuGluGlyGluProSerIleSer 172
654 GCTGGAACTGGGTGGGCACAATTGCGAGCTGTGACGACTATGGGCGCGG 703
173ArgAlaAlaIleThrPheSerT 180
704 GGGATTGAGAAATTCGGAGAGAACTGAGGAAGGGATATCTGCATCGA 753
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
: : : : :
754 CTTCACTGAACTCATCTCCAGTACTCTGATCAGGAAGAGATCCAGCATG 803
196ArgGluGluSerArgIleLeuLeuG 204
804 TGGTAGAGGTGATCAAAATTCACGGCCCAAGTCACTCGTGGTTTCTCC 853
204 lnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
: : : : :
854 ATGGCCAGATCTTGAGCCCTCATCAGGAGATGTCGCGCGCAA... 900
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArg 237
: : : : :
901TATCAGGGCAAGATCTGCTGGCCAGCGAGGCTGGGCCAGCT 944
237 gGlyProAsnGlnGlyProArgGlyLeuGlyHisSerT 250
945 CTTCTCTGATGCCCATGCTCAGTACTTCCACGTGGTGGCGGCACCAT 994
250 rpArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysrtpSer 266
: : : : :
995 GGATTCGCTCTGAAGCTGGGCA 1017
267 ProProTyrLeuGluCy 272
1018 GATCCAGGCTTCGGGAATTCCTGAAGAAGTCCATCCAGGAAGTCTG 1067
272 sGluAsnGlySerTyrLysProGlyTrpLeuValThr 284
1068 TCACAAATGGTTTCCCAAGGAGTTTGGGAAGAACAATTTAACTGCCAC 1117
285 LeuSerSerAlaIleTyrGlyLeuGlnPro 294
1118 CTCCAAGAGGTGCAAAAGGACCTTACTCTGGACACCTTCTGAGAGG 1167
295AsnLeuValProGluPheArgg 302
1168 TCACGAAGAAGTGGCAGAGTTTAGCAACAGCTCCGACAGCTTCCGAC 1217
302 lyValMetLysValAspIleAsnLeuGlnLysVal 313
1218 CCCTCTGACAGGGGATGAGAACATCAGCAGTGTGAGACCCCTTACATA 1267
313 313
1268 GATTACAGCATTTACGGATATCTTACAAATGTGTACTTAGCAGTCTACTC 1317
314AspIleAspGlnCys 318
1318 CATTCGCCACGCTTTCGAAGATATATACCTGTACCTGGAGAGGCG 1367
319SerSerAspGlyTrpPhe 324
1368 TCTTCACCAATGGCTCTCTGTGCAGACATCAAGAAAAGTTGAGCGGTGGCAG 1417
325 SerGlyThrHisLysCysHisLeuAsnAsnSer 335
1418 GTCTTGAAGCACCTACCGCATCTAACTTTACAAACAATATGGGGAGCA 1467
336GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
1468 GGTGACCTTTGATGAGTGTGGTGACCTGGTGGGAACCTATTCCATCATCA 1517
348Gly 348
1518 ACTGCGACCTCTCCCGAGGAGATGCTCCATCGTGTAAAGGAAGTCGGG 1567
349 AlaTyrGluCysIleCysLysAlaGlyPheTyrHisProGl 362
1568 TATTACAAGCTATGCCAAGAGGAGAAAGACTCTTCATCATCAACGAGGA 1617
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis.... 377
1618 GAAATCTCTGGAGTGTCTCCAGGAGGTGCCCTTCTCCAACATGCA 1667
378IleSerGlySerThrLysAspValSerGluGlu..... 388
1668 GCCGAGACTGCTGGCAGGGACCAAGGAAGGATCATTTGAGGGGAGGCC 1717
389AlaTyrValCysLeuProCysArgGluGly 398
1718 ACCTCTCTTTGAGTGTGGAGTGTCTGATGGGAGTATAGTGTATGA 1767
399CysProPheCysAlaAspAsp 405
1768 GACAGATGCCAGTGCCTGTAACAAGTCCCGACAGACTCTGTGGTCCAATG 1817
406SerProCysPheValGlnGluAspLysTyrLeuArg 417
1818 AGAACACACCTCTCGGAGGCCATTCCTGCTCGGTATCTTGAGTGGAGT 1867
418LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
1868 GACATAGATCTATCATAGCCATCGCTTTTCTTCCCTGGGCATCCTCGT 1917
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
1918 GACGCTGTTGTACCCCTCATCTCGTTCTGTACCGGGACACACCCGTGG 1967
447 leArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeu 463
1968 TCAATCTCCAGTAGGAGCTGCTATATCATCTCTGGCTGGTATTTC 2017
464 LeuLeuTyrPheProValIleLeuTyrPheGluProSerThrPheAr 480
2018 CTGCGCTATGTGCCCTTTCACCCCTCATCGCCAAACCTACTACACATC 2067
480 gCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrG 497
2068 CTGCTACCTCCAGCGCCCTCCTAGTTGGCTCTCTCTGCCATGTGCTACT 2117
497 lyThrValThrLeuLysLeuHisArgValLeuLysVal 509
2118 CTGCTTTAGTACCAAAACCAATCGTATTGACCGCATCTCTGGCTGGCAGC 2167
510PheLeuSerArgThrAlaGl 516
2168 AAGAAGAAGATCTGCACCCGGAAGCCCGAGATTTCATGACGCTGGGCCCA 2217
516 nArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAlaValI 533

```

1      |||
2218  ATGATCATAGCCCTCCATTCTGATTAGTGACAGCTAACACTAGTGGTGA 2267
533  leLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerSerValCys 549
2268  CCTTGATCATC..... 2278
550  GlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAs 566
2279  ....ATGAGAGCTCCCATGCGCCATTTTGTCTTACCAGATATCAAGGA 2322
566  pHISLeuIlePheAsnMetCysLeuIleAspArgTTPAspTyrMetThrA 583
2323  A.....GTCTACTATCTGCAATACCAGCAACCTGGGTGTAGTGGCCC 2366
583  laValAla...GlupheLeuPheLeuLeuTrpGlyValTyrLeuCysTyr 598
2367  CTGTGGGTTACAATGGACTCCTCATCATGAGCTGCTACTACTATGCCTTC 2416
599  AlavalArgThrValProSerAlaPheHisGluProHrGlyTyrMetAlaVa 615
2417  AAGACCGCAACGTGGCGGCAACTTCAATCAGGCTAAATACATCGCCTT 2466
615  lAlaValHisAsnGluLeuIleSerAlaIlePheHisThrIleArgp 632
2467  CACATGTACACTACCTGCATCATCTGGCTGCTTCTGTTCCCATTTACT 2516
632  heValLeuAlaSerArgLeuGlnSerAspTrpMetLeuMetLeuTyrPhe 648
2517  TT.....GGGAGCAACTACAGATCATCATCTGC 2548
649  AlAHISThrHisLeuThrValThrValThrIleGlyLeuLeuIlePr 665
2549  TTCGCGGTGAGCCTCAGTGTGACGGTGGCCCTGGGTGCATGTTTACTCC 2598
665  oLys.....PheSerHisSerAsnAsnProArgAspAspI 678
2599  GAGATGTACATCATCTGCCAACTGAGAGGACGTCCGAGTGCCT 2648
678  leAlaThrGluAlaTyrGluAspGluLeuAspMetGly..... 690
2649  TCAGGACC.....TCTGATGTTTCCGCATGCACGTGGGTGATGGC 2689
691  ....ArgSerGlySerTyrLeuAsnSerSerIleAsnSerAl 703
2690  AAATGCGCGTGGCGCTCCAACTCCCTCAAC..... 2722
703  aTrpSerGluHisSerLeuAspProGluAspIleArgAspGluLeuLysL 720
2722  ..... 2722
720  ysLeuTyAlaGlnLeuGluIleTyrLysArgLysLysMetIleThrAsn 736
2723  .....ATTTCGGAGAGAGAGCCGGGACAGG 2752
737  AsnProHisLeuGlnLysLysArgCysSer..... 746
2753  AATGCCAATTTCTAACGCCAAGTCTGTGTCATGTGTCGAACAGGTGGAAG 2802
747  LysLysGlyLeuGlyArgSerIleMetArgArgIleThrGluIleProG 763
2803  ACAGGCCCAAGGAGCAGCAGTGTGGCAGCGCTCTCTGTGCACGTGA 2852
763  luthrValSerArgGlnCysSerLysGluAspLysGluGlyAlaAspHis 779
2853  AGACCAACGAGAGCGCTGTAAACAA..... 2878
780  GlyThrAlaLysGlyThrAlaLeuIleArgLysAsnProProGluSerSe 796
2879  .....ACAGCGTAATCAAA.....CCCTCACTAA 2904
796  rGlyAsnThrGlyLysSerLysGluGluThrLeuLysAsnArgValPheS 813
      |||..... |||.....

```

```

2905  AAGTTACCAAGGCTCTGGCAAGACCTGACCTTTTCAGAT..... 2944
813  erLeuLysLysSerHisSerThrTyrAspHisValArgAspGlnThrGlu 829
2944  ..... 2944
830  GluSerSerSerLeuProThrGluSerGlnGluGluThrThrGluAs 846
2944  ..... 2944
846  nSerThrLeuGluSerLeuSerGlyLysLysLeuThrGlnLysLeuLysG 863
2945  .....GCCAGCACCAAGACCTTTACAATGTGGAG 2975
863  luAspSerGluAlaGluSerThrGluSerValProLeuValCysLysSer 879
2976  AAGAGGACAATACCCCTTCTGCTCACITTCAGCCCTCCAGCAGCCCTTCT 3025
880  AlaSerAlaHisAsnLeuSerSerGluLysLysThrGlyHisProArgTh 896
3026  ATGGTGTGTGACCGACGCGCGGCCACCGTGGCCACACACACCTCTGCC 3075
896  rSerMetLeuGlnLysSerLeuSerValIleAlaSerAlaLysGluLysT 913
3076  ACCCATCTG.....ACCGCAGAGAGACCC 3101
913  hrLeuGlyLeuAlaGlyLysThrGlnThrAlaGlyValGluGluArgThr 929
3102  CCCTGTTCCTGGTGTATCCGTATCCATCCCAAGGCTTG..... 3139
930  LysSerGlnLysProLeuProLysAspLysAspGluThrAsnArgAsnHisSe 946
3140  .....CCTCTCTCTCTCCCGCAGCAG..... 3163
946  rAsnSerAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsnp 963
3164  .....CCACAGCAGCGCCGCC 3179
963  roAlaGluGluProArgLysProGlnLysSerGlyIleMetLysGlnGln 979
3180  CTCAGCAGCCCCCGCAGCAGCCCCAAG.....TCCCTGTATGGACCACTG 3223
980  ArgValAsnProThrThrAlaAsnSerAspLeu..... 990
3224  CAAGCGCTAGTACCAACTTCGGTTCGGGATTCAGATTTCATGCGGT 3273
991  .....AsnProGlyThrThrGlnMetLysAspAsnPheAspIleGlyG 1005
3274  GCTGGCAGGCGCGGGACACCA.....GGAA 3299
1005  luValCysProTrpGluValTyrAspLeuThrProGlyProValProSer 1021
3300  AC.....AGCCTGGCTCTCTGTACCGCCCGCCCTCCCGCG 3337
1022  Glu.....SerLysValGlnLysHisValSerIleValAlaSerGluMe 1036
3338  CAACACCTGCAGATGCTGCCCTGCACCTGAGCACCTTCAGGAGGAGTC 3387
1036  tGluLysAsnProThrPheSerLeuLysGluLysSerHisHisLysProL 1053
3388  CATCTCCCTCTCTGGGAGGACATCGATGATGACAGT..... 3424
1053  ysAlaAlaGluValCysGlnSerAsnGlnLysArgIleAspLysAla 1069
3425  .....GAGACATTCAGCTCCTG 3442
1070  GluValCysLeuTrpGluSerGlnGly.....GlnSerIleLeuGl 1083
3443  CAGGAGTTCGTACGACGCGAGGGAACACCAAGAAAGATGAATTCGA 3492
1083  uAspGluLysLeuLeuIleSerLysThrProValLeuProGluAlaL 1100
3493  AGAGGAGGAGGACCTGCCACAGCCAGCTGACCCCTGAGGATTCCTC 3542

```


1502 TTCAGTGTGCGCTACCTGACACCTCTTGAAACCCCACTTTAAGAA 155

[illegible]

```

2443 CAAGAAGAAGATCTGACCCGGAAGCCAGATTTCATGAGCGTTGGGCC 2492
||||:|||||
516 InArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAlaVal 532
||| |||:|||||
2493 AAGTGATCATCCCTCCATTCGATTAGTACAGCTACACACTAGTGGTG 2542
||| |||:|||||
533 IleLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerSerValCy 549
||||:|||||
2543 ACCTTGATCATC 2554
||||:|||||
549 sGlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSera 566
||||:|||||
2555 .....ATGAGACCTCCCATCCCACTTTGCTACCCGAGATCAAGG 2597
||||:|||||
566 sPheHisLeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrMetThr 582
||| |||:|||||
2598 AA.....GTCTACCTTATCTGCAATACCAAGCAACCTGGGTAGTGCC 2641
||||:|||||
583 AlaValAla...GluPheLeuPheLeuLeuTrpGlyValTyrLeuCysTy 598
||||:|||||
2642 CCTGTGGTTTACAATGGAGCTCTCATCATGAGCTGTACCTACTATGCGCT 2691
||||:|||||
598 rAlaValArgThrValProSerAlaPheHisGluProArgTyrMetAlaVal 615
||||:|||||
2692 CAAGACCCGCAACGTCGCGGCAACTTCAATGAGGCTAAATACATCGCGCT 2741
||||:|||||
615 alAlaValHisAsnGluLeuIleIleSerAlaIlePheHisThrIleArg 631
||||:|||||
2742 TCACCATGTACACTACCTGCATCATCTGGCTGGCTTTCGTCCCATTTAC 2791
||||:|||||
632 PheValLeuAlaSerArgLeuGlnSerAspTrpMetLeuMetLeuTyrPh 648
||| |||:|||||
2792 TTT.....GGGAGCAACTACAAGATCATCATACCTG 2823
||||:|||||
648 eaIaHisThrHisLeuThrValThrValThrIleGlyLeuLeuIleP 665
||||:|||||
2824 CTTCCGGGTGACCTTCAGTGTGACCGTGGCTGGGTGCATGTTTACTC 2873
||||:|||||
665 rOLys.....PheSerHisSerSerAsnAsnProArgAsp 677
||||:|||||
2874 CGAAGATGTACATCATCTGCCAACCTGAGAGGAAGCTCGGCAAGTCC 2923
||||:|||||
678 IleAlaThrGluAlaTyrGluAspGluLeuAspMetGly..... 690
||||:|||||
2924 TTCAGGACC.....TCTGATGTTTCCGCATGCACGTCGGTGATGG 2964
||||:|||||
691 .....ArgSerGlySerTyrLeuAsnSerSerIleAsnSera 703
||||:|||||
2965 CAAACTGGCGTCCGCTCCACACCTTCTCAAC..... 2998
||||:|||||
703 IaTrpSerGluHisSerLeuAspProGluAspIleArgAspGluLeuLys 719
||||:|||||
2998 ..... 2998
720 LysLeuTyrAlaGlnLeuGluIleTyrLysArgLysLysMetIleThrAs 736
||||:|||||
2999 .....ATTTTCGGAGAAAGACCGCGGCGCAGG 3027
||||:|||||
736 naSnProHisLeuGlnLysLysArgCysSer..... 746
||||:|||||
3028 GAATCCCAATTCATCGGCAAGTCTGTCTCATGGTCTGAACACAGTGGAA 3077
||||:|||||
747 ..LysLysGlyLeuGlyArgSerIleMetArgArgIleThrGluIlePro 762
||| |||:|||||
3078 GACAGCGCCCAAGGACAGCAGCTGTGCGAGCGCTCTCTCTGACACGTG 3127
||||:|||||
763 GluThrValSerArgGlnCysSerLysLysGluAspLysGluGlyAlaAspHi 779
||||:|||||
3128 AAGACCAACGAGACGCGCTGTAAACAA..... 3154
||||:|||||
779 sGlyThrAlaLysGlyThrAlaLeuIleArgLysAsnProProGluSers 796
||||:|||||

```

```

3155 .....ACAGCCGTAAATCAAA.....CCCTCACTA 3179
796 erGlyAsnThrGlyLysSerLysGluGluThrLeuLysAsnArgValPhe 812
||||:|||||
3180 AAAGTTACCAAGCTCTGGCAGAGCCCTGACCTTTTCAGAT..... 3220
813 SerLeuLysLysSerHisSerThrTyrAspHisValArgAspGlnThrGl 829
3220 ..... 3220
829 uGluSerSerSerLeuProThrGluSerGlnGluGluThrThrGluA 846
3220 ..... 3220
846 snSerThrLeuGluSerLeuSerGlyLysLeuThrGlnLysLeuLys 862
3221 .....GCCAGCACCAGACCCCTTTACATGTGGAA 3250
863 GluAspSerGluAlaGluSerThrGluSerValProLeuValCysLysSe 879
||||:|||||
3251 GAAAGAGCAATAACCCCTCTGCTCACTTCAGCCCTCCAGCAGCCCTTC 3300
879 rAlaSerAlaHisAsnLeuSerSerGluLysLysThrGlyHisProArgT 896
||||:|||||
3301 TATGTGTGTGCACCGCAGCGGGCCACCGTGGCCACACACCTCTGC 3350
896 hrSerMetLeuGlnLysSerLeuSerValIleAlaSerAlaLysGluLys 912
||||:|||||
3351 CACCCCATCTG.....ACCGAGAAGAGACC 3376
913 ThrLeuGlyLeuAlaGlyLysThrGlnThrAlaGlyValGluGluArgTh 929
||||:|||||
3377 CCCCTGTCTCTGGCTGATTCCGTCATCCCAAGGGCTTG..... 3415
929 rLysSerGlnLysProLeuProLysAspLysGluThrAsnArgAsnHis 946
||||:|||||
3416 .....CCTCCTCCTCTCCGCGCAGCAGCAG..... 3439
946 erAsnSerAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsn 962
3440 .....CCACAGCAGCGCGCC 3454
963 ProAlaGluGluProArgLysProGlnLysSerGlyIleMetLysGlnGl 979
||||:|||||
3455 CCTCAGCAGCGCGCGCAGCAGCAG.....TCCCTGATGGACCACT 3498
979 nArgValAsnProThrThrAlaAsnSerAspLeu..... 990
3499 GCAAGGCGTAGTCACCAACTTCGGTTCGGGATTCCAGATTTCCATGCGG 3548
991 .....AsnProGlyThrThrGlnMetLysAspAsnPheAspIleGly 1004
||||:|||||
3549 TGCTGGCAGCGCGGGGACCA.....GGA 3574
1005 GluValCysProTrpGluValTyrAspLeuThrProGlyProValProSe 1021
||||:|||||
3575 AAC.....AGCCTGCGCTCTCTGTACCGCGCGCGCTCCGCC 3612
1021 rGlu.....SerLysValGlnLysHisValSerIleValAlaSerGlu 1036
||||:|||||
3613 GCAACACCTGCAGATGTCGCCCTGCACCTGAGCACCTTCACGAGGAGT 3662
1036 etGluLysAsnProThrPheSerLeuLysGluLysSerHisHisLysPro 1052
||||:|||||
3663 CCATCTCCCTCCTGGGAGGACATCGATGATGACAGT..... 3700
1053 LysAlaAlaGluValCysGlnSerAsnGlnLysArgIleAspLysAl 1069
||||:|||||
3701 .....GAGAGATTCAAGCTCCT 3717
1069 aGluValCysLeuTrpGluSerGlnGly.....GlnSerIleLeuG 1083
||||:|||||
3718 CGAGGAGTTCGTGTACGAGCGCGAAGGGAACACCGAAGAGATGAATTGG 3767

```


1502 TTTCCAGTGTGGCTACCTGGACAGCTCTTTGGAAAAACCCCAACTTTAAGAA 1551
297ValProGluPheArgGlyValMetLysValAspIleAsnLeuGln 311
1552 AGTGTGCACAGANAATGAAGCTTGAAGAAAACACTATGTCCAGGACAGCA 1601
312 LysValAspIleAspGlnCysSerSerAspGlyTrpPheSerGlyThrHi 328
1602 AAATGGGATTTGTTCATCAATGCCATCTATGCCATGCCACATGGGCTGCAG 1651
328 sLysCys..His.....LeuAsnAsnSerGluCysMe 338
1652 AACATGCACCATGCTGTGTCTCCGGCCCATGTGGCCCTGTGTGATGCTAT 1701
338 t....ProIleLysGly.....LeuGlyPheValLeu..... 347
1702 GAACCCATGTATGGCAGGAAGCTCTGTGATTTCTTCATCAATCTCTTT 1751
347 347
1752 TTTCTGGAGTGTCTGGAGGAGGTGTGTCTCGATGAGAAGGGGATGCT 1801
348 ...GlyAlaTyrGluCysIle..... 353
1802 CCGGGAAGGTATGACATTATGAATCTGCAGTACACAGAAGCTAAATCGCTA 1851
354CysLysAlaGlyPheTyrHisProGlyValLeuProValAsnA 368
1852 TGACATGTCCAGTGGGACCTGCATGAGGAGTGTGATATATGATG 1901
368 snPheArgArgArgGlyProAspGlnHisIleSerGlySerThrLysAsp 384
1902 AATCAAAATCCAG.....ATGACAAAAGCGGAATGTACGATCT 1942
385 ValSerGluGlu..... 388
1943 GTGTGCAGTGGACCTTGTCTTAAGGGTCTAGATTAAGGTCTATACGGAAAGG 1992
389AlaTyrValCysLeuProCysArgGluGly..... 398
1993 AGAAGTGAAGTCTGCTGGATCTGCAGCGCTGCAGAGAGATGATGTTG 2042
399CysProPheCys..... 402
2043 TGCAGGAGGAGTTTCACTGTCAGAGCTGTGACCTGGGTGTGGGCCCAAC 2092
403 AlaAspAspSerProCysPheValGlnGluAspLysTyrLeuArg..... 417
2093 GCAGAGCTCACAGGCTGTGAGGCCCATCTCTGCTCCGTTATCTGTAGTGGAG 2142
418LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuL 430
2143 TGACATAGATCTATCATAGCATCGCTTTCTTCTGCTGGCATCTCTCG 2192
430 euAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSer 446
2193 TGACGCTGTTTGTACCTTCATCTCTGTTCTGTACCGGACACACCCCGTG 2242
447 IleArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerIe 463
2243 GTCAATCTCCAGTAGGAGCTCTGCTATATCATFTCTGGCTGATTTT 2292
463 uLeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheA 480
2293 CCTGGCTATGTGGCTTTTACCTCTCATCGCCAAACCTACTACCAT 2342
480 rgCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyr 496
2343 CTGTCTACCTCCAGCGCTCTCTAGTTGGCTCTCTCTGCTGATGTGCTAC 2392
497 GlyThrValThrLeuLysHisArgValLeuLysVal..... 509
..... 509

2393 TCTGTTTATAGTGACCAAAACCAATCGTATTGACGCATCTCTGGCTGGCAG 2442
510PheLeuSerArgThrAlaG 516
2443 CAAGAAGAGATCTGCACCCGGAAGCCAGATTCATGAGCGCTTGGGCC 2492
516 InArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAlaVal 532
2493 AAGTATCATAGCTCCATTCATGTAGTATAGTACAGCTAACACTAGTGGTG 2542
533 IleLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerSerValCy 549
2543 ACCTTGATCATC..... 2554
549 sGlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerA 566
2555ATGGAGCTCCCATGCCCATTTTGTCTTACCAGGATATCAAGG 2597
566 sPheIleLeuPheAsnMetCysLeuIleAspArgTrpAspTyrMetThr 582
2598 AA.....GTCTACCTTATCTGCAATACCAGCAACCTGGGTGTAGTGCC 2641
593 AlavalAla....GluPheLeuPheLeuLeuTrpGlyValTyrLeuCysty 598
2642 CTGTGGGTACATGAGTCTCTCATGAGCTGTACTATGCTT 2691
598 rAlaValArgThrValProSerAlaPheHisGluProArgTyrMetAlav 615
2692 CAAGACCCGCAACGTGCGGCAACTTCAATGAGGCTAAATACATCGCCT 2741
615 alAlaValHisAsnGluLeuIleIleSerAlaIlePheHisThrIleArg 631
2742 TCACCATGTACACTACCTGCATCATCTGGCTTGTCTTCCCATTTAC 2791
632 PheValLeuAlaSerArgLeuGlnSerAspTrpMetLeuMetLeuTyrPh 648
2792 TTT.....GGGAGCAACTACAAAGATCATCATACCTG 2823
648 eAlaHisThrHisLeuThrValThrValThrIleGlyLeuLeuLeuIleP 665
2824 CTTGCGGTGAGCTCAGTGTGACGGTGGCTGGGTGCATGTTTACTC 2873
665 rOLys.....PheSerHisSerSerAsnAsnProArgAspAsp 677
2874 CGAAGATGTACATCATCTGCCAAACCTGAGAGCACTCCGACGTGCC 2923
678 IleAlaThrGluAlaTyrGluAspGluLeuAspMetGly..... 690
2924 TTCAGACC.....TCTGATGTTTCTCGCATGTCACGTCGGGTGATGG 2964
691ArgSerGlySerTyrLeuAsnSerSerIleAsnSerA 703
2965 CAACACTGCGTGGCTCCACACCTTCTCTCAAC..... 2998
703 latrPserGluHisSerLeuAspProGluAspIleArgAspGluLeuLys 719
2998 2998
720 LysLeuTyrAlaGlnLeuGluIleTyrLysArgLysLysMetIleThrAs 736
2999ATTTCGGAGAAGAAAGCCCGGGCAGG 3027
736 nAsnProHisLeuGlnLysArgCysSer..... 746
3028 GAATGCCAATTTCTAACGGCACTGTGTCTCATGGTCTGAACCAAGGTGAA 3077
747 ..LysLysGlyLeuGlyArgSerIleMetArgIleThrGluIlePro 762
3078 GACAGCGCCCAAGGACAGCACGTGTGGCAGCGCTCTCTGTGTCAGCTG 3127
763 GluThrValSerArgGlnCysSerLysGluAspLysGluGlyAlaAspHi 779
3128 AAGACCAACGACGCGCTGTAAACCA..... 3154


```
3701 .....GAGAGATTCAAGCTCCT 3717
1069 aGluValCysLeuTrpGluSerGlnGly.....GlnSerIleLeuG 1083
3718 GCAGGAGTTCGTACAGCGGAGGAGCAACCCGAGAGATGATTTGG 3767
1083 luAspGluLysLeuIleSerLysThrProValLeuProGluAla 1099
3768 AAGAGGAGGAGGACCTGCCACAGCCAGCAAGCTGACCCCTGAGGATTCT 3817
1100 LysGluLysAsnGlyClyGlnProArgAlaAlaAsnValCysAlaGlyG 1116
3818 CCGGCCCTCAGCCCTCTTCCTCTTCGAGATTCCGTGGCCTTGGCAG 3867
1116 nSerGluGluLeuProPro.....LysAlaValAlaSerLysThrGluA 1131
3868 CTCAGTCCCGCAGTTCCTCCGTCATCTGAGTCGGTCTCTCCACCCCTCCAA 3917
1131 snGluAsnLeuAsnGlnIleGlyHisGlnGluLysLysThrSerSer 1147
3918 ATGTAACCTACGCTCTGTCTCATCTGAGGAGTCTACAAAGCAAGCTCTTCC 3967
```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-465-157-1

```
seq_documentation_block:
; Sequence 1, Application US/08465157
; Patent No. 5869609
; GENERAL INFORMATION:
; APPLICANT: Muvihiill, Eileen R
; APPLICANT: Hagen, Frederick S
; APPLICANT: Houamed, Khaled M
; APPLICANT: Almers, Wolfhard
; TITLE OF INVENTION: G PROTEIN COUPLED GLUTAMATE
; TITLE OF INVENTION: RECEPTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: Steuart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,157
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/041,538
; FILING DATE:
; APPLICATION NUMBER: US/07/648,481
; FILING DATE:
; APPLICATION NUMBER: US 07/626,806
; FILING DATE: 12-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 13952-6-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 206-623-6793
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4300 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
```

```
; ORIGINAL SOURCE:
; ORGANISM: Rattus norvegicus
; TISSUE TYPE: Cerebellum
; IMMEDIATE SOURCE:
; CLONE: 45-A
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 377..3973
; OTHER INFORMATION:
; US-08-465-157-1

alignment_scores:
Quality: 232.00 Length: 1234
Ratio: 0.480 Gaps: 51
Percent Similarity: 39.141 Percent Identity: 19.044

alignment_block:
US-09-775-181-2 x US-08-465-157-1 ..
Align seg 1/1 to: US-08-465-157-1 from: 1 to: 4300

153 AspLeuAspTrpTrpTyrGlnAlaLeuValTrpSerLeuGluGlyGluPr 169
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
853 GATCGGCCCTGGCTCCGACTCTGTGGCCATTCAAGTCCAGAAATCTTCTCC 902
169 oSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeu..... 183
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
903 A.....GCTGTTCGACATCCACAGATCCGCTATTCTTCTGCC 937
184 .....SerAlaProAlaProGlnValPheLeuGlnAlaThrArgGlu... 197
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
938 ACAAGCATAGACCTGAGTCACAAACTTTGTACAAATCTTCTGAGGGT 987
198 .....GluSerArgIle 201
988 GGTCCCTTCTGACACTTTTCAGGAGGCGGATGCTCGACATAGTCAAGC 1037
201 eLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrL 218
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1038 GTTAACTGGACCTATGTCTCAGCAGTCCACACAGAGGGAA..... 1080
218 euGluThrGluTrpPheHisGlyLeuArgArgLysTrp..... 230
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1081 .....TTACGGGAGAGTGGAAATGGATGCTTT 1107
231 .....ArgProHisLeuHisArgArg....GlyPr 239
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1108 CAAAGAACTGGCTGCCAGGAAGGCT...CTGCATCGCACACTCGGACA 1154
239 oAsn..... 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1155 AAATCTACAGCAATGCTGGCGAGAGAGCTTTGACCGGCTCTGCGTAA 1204
241 .....GlnGlyProArgGlyLeuGlyHisSerTrpArg 251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1205 CTCGGGAGGGCTTCCCAAGGC...CAGGGTTGGGTCTGCTCTCTCGGA 1251
252 ArgLysAspGlyLeuGlyCysLysSerHis..PheLysTrp..... 265
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1252 GGGCATGACAGTGGCGGGCTTACTGATGCCATGCCGCCCTGGCGCTCG 1301
256 .....SerProTyrLeuGluCysGluAsnGlySer..... 276
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1302 TGGGCGAGTTCTACTCAVTTGGAGTGATGGGACAGACAGATGAA 1351
277 .....TyrLysPro..... 279
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1352 GTCATCGAAGGCTATGAGTGGAGGCAAGGAGGATCACAATAAAGCT 1401
280 .....Gly 280
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1402 TCAGTCTCCAGAGTCAAGTCAATTTGATGACTTCTTGAAGCTGAGGC 1451
```



```

3078 GACAGCGCCGACAGGACAGCAGTGTGACGCGCTCTGTGACAGTG 3127
763 GluThrValSerArgGlnCysSerLysGluAspLysGluGlyAlaAspH1 779
3128 AAGACCAACGAGAGCGCTGTACCAAA..... 3154
779 sGlyThrAlaLysGlyThrAlaLeuIleArgLysAsnProProGluSerS 796
3155 .....ACAGCGGTAAATCAA.....CCCTCACATA 3179
796 erGlyAsnThrGlyLysSerLysGluGluThrLeuLysAsnArgValPhe 812
3180 AAGTTTACCAAGCGCTGTGCAAGAGCGCTGACCTTTTCAGAT..... 3220
813 SerLeuLysLysSerHisSerThrTyrAspHisValArgAspGlnThrGl 829
3220 ..... 3220
829 uGluSerSerSerLeuProThrGluSerGlnGluGluThrThrGluA 846
3220 ..... 3220
846 snSerThrLeuGluSerLeuSerGlyLysLysLeuThrGlnLysLeuLys 862
3221 .....GCCAGCACCAAGAGCCCTTTACAAATGTGGAA 3250
863 GluAspSerGluAlaGluSerThrGluSerValProLeuValCysLysSe 879
3251 GAAGAGGACAAATACCCCTCTCTCTACCTACCTCCAGCAGCCCTTC 3300
879 rLaserAlaHisAsnLeuSerSerGluLysLysThrGlyHisProArgT 896
3301 TATGTTGTGTGACCCAGCGGGGCCACCGGTGGCCACCACACCTCTGC 3350
896 hrSerMetLeuGlnLysSerLeuSerValIleAlaSerAlaLysGluLys 912
3351 CACCCCATCTG.....ACCGCAGAGAGAGACC 3376
913 ThrLeuGlyLeuAlaGlyLysThrGlnThrAlaGlyValGluGluArgTh 929
3377 CCCCTGTTCTTGCTGATTCGCTCATCCCAAGGCGTTG..... 3415
929 rLysSerGlnLysProLeuProLysAspLysGluThrAsnArgAsnHisS 946
3416 .....CCTCCTCTCTCCCGCAGCAG..... 3439
946 erAsnSerAspAsnThrGluThrLysAspProAlaProGluAsnSerAsn 962
3440 .....CCACAGCAGCGCCGCC 3454
963 ProAlaGluGluProArgLysProGlnLysSerGlyIleMetLysGlnGl 979
3455 CTTACAGAGCCCCCGCAGCAGCCCAAG.....TCCCTGATGGACCACT 3498
979 nArgValAsnProThrThrAlaAsnSerAspLeu..... 990
3499 GCAAGGCGTAGTCAACCACTTCGGTTCGGGATTCACAGATTCCATCGCG 3548
991 .....AsnProGlyThrThrGlnMetLysAspAsnPheAspIleGly 1004
3549 TGCTGGCAGGCGCGGGGACACCA.....GGA 3574
1005 GluValCysProTrpGluValTyrAspLeuThrProGlyProValProSe 1021
3575 AAC.....AGCTGGCTCTCTGTACCGCGCCCGCCCTCCGCC 3612
1021 rGlu.....SerLysValGlnLysHisValSerIleValAlaSerGlu 1036
3613 GCAACACCTGCAGATGCTGCGCCCTGCACCTGAGCACCCTTCAGGAGGAGT 3662
1036 etGluLysAsnProThrPheSerLeuLysGluLysSerHisHisLysPro 1052
...
3663 CCATCTCCCTCTCTGGGAGGACATCGATGATGACAGT..... 3700
1053 LysAlaAlaGluValCysGlnGlnSerAsnGlnLysArgIleAspLysAl 1069
3701 .....GAGAGATTCAAGCTCCT 3717
1069 aGluValCysLeuTrpGluSerGlnGly.....GlnSerIleLeuG 1083
3718 GCAGGAGTTCGTGTACGAGCGGAAGGAACACCAAGAAAGATGAATTGG 3767
1083 luAspGluLysLeuLeuIleSerLysThrProValLeuProGluArgAla 1099
3768 AAGAGGAGGAGGACCTGCCACGACGACGACCTGACCCCTGAGGATCT 3817
1100 LysGluGluAsnGlyGlyGlnProArgAlaAlaAsnValCysAlaGlyGl 1116
3818 CCTGCCCTCAGCCCTCTCTCTTCCGACGATTCCTGGGCTCTGGCAG 3867
1116 nSerGluGluLeuProPro.....LysAlaValAlaSerLysThrGluA 1131
3868 CTCAGTGCCTCAGTTCCTCTATCTGATCGTCTCTGCACCCCTCCAA 3917
1131 snGluAsnLeuAsnGlnIleGlyHisGlnGluLysLysThrSerSerSer 1147
3918 ATGTAACCTACGCTCTCTCTATCTGAGGGGACTACAAAGCAAGCTCTTCC 3967
seq_name: /cgn2_6/ptodata/2/1na/PCTUS_COMB.seq:PCT-US91-09422-1
seq_documentation_block:
; Sequence 1, Application PC/TUS9109422
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R.
; APPLICANT: Hagen, Frederick S.
; APPLICANT: Houamed, Khaled M.
; APPLICANT: Almers, Wolfhard
; TITLE OF INVENTION: G PROTEIN-COUPLED GLUTAMATE RECEPTORS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: One Market plaza, Steuart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/09422
; FILING DATE: 19911212
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,007
; FILING DATE: 18-MAR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/648,481
; FILING DATE: 30-JAN-1991
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/626,806
; FILING DATE: 12-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 13952-6PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 467-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4300 base pairs
; TYPE: NUCLEIC ACID
```



```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: 45-A
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 377..3973
; PCT-US91-09422-1
```

```
alignment_scores:
  Quality: 232.00      Length: 1234
  Ratio: 0.480        Gaps: 51
  Percent Similarity: 39.141  Percent Identity: 19.044
```

alignment_block:

US-09-775-181-2 x PCT-US91-09422-1 ..

Align seg 1/1 to: PCT-US91-09422-1 from: 1 to: 4300

```
153 AspLeuAspTrpTyrGlnAlaLeuValTrpSerLeuLeuGluGlyGluPr 169
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
853 GATCGGCCCTGCTCCAGCTCTGCGCCATTCAAGTCCAGAACTTCTCC 902
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169 oSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeu..... 183
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
903 A.....GCTGTCGACATCCACAGATCGCTATTCTGCC 937
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 .....SerAlaProAlaProGlnValPheLeuGlnAlaThrArgGlu... 197
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
938 ACAAGCATAGACCTGAGTGACAAAACTTGTACAAATACATTCTCTGAGGGT 987
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
198 .....GCTGTCGACATCCACAGATCGCTATTCTGCC 937
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
988 GGTCCCTTCGACATTTCGAGCAAGGGCGATGCTCGACATAGTCAAGC 1037
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 eLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrL 218
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1038 GTTACAACTGGACCTATGCTCAGCAGTCCACAGAGGAA..... 1080
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 euGluThrGluTrpPheHisGlyLeuArgArgLysTrp..... 230
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1081 .....TTACGGCGAGAGTGGAAATGATGCTTT 1107
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
231 .....ArgProHisLeuHisArgArg...GlyPr 239
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1108 CAAAGAACTGGCTGCCAGGAGGCT...CTGCATCGCACATCGGACA 1154
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
239 oAsn..... 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1155 AAATCTACAGCAATGCTGGCGAGAAGAGCTTTGACCGGCTCTCTCGGTAA 1204
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 .....GlnGlyProArgGlyLeuGlyHisSerTrpArg 251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1205 CTCGGGAGCGGCTTCCCAAGC...CAGGTTGTGCTGCTCTCGCA 1251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
252 ArgLysAspGlyLeuGlyAspLysSerHis...PheLysTrp..... 265
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1252 GGCATGACACAGTCCGGGCTTACTGAGTGCATCGCCCGCTGGCGTCG 1301
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 .....SerProTrpTyrLeuGlyCysGluAsnGlySer..... 276
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1302 TGGGCGAGTCTCACTCATTTGGAAGTGTGATGGATGGCGACAGAGATGA 1351
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
277 .....TyrLysPro..... 279
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1352 GTCATCAAGGCTATGAGTGGGAAGCAAGGAGGAGGATCAATAAAGCT 1401
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
280 ..... 280
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1402 TCAGTCTCCAGAGGTGAGTTCATTGTTGATGACTACTTCTCTGAAGCTGAGGC 1451
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
281 TrpLeuValThr.....LeuSe 286
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1452 TGGACACCAACACAGGAATCCTTTGTTCCCTGAGTTCTGGCAACATCGC 1501
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 rSerAlaIleTyrGlyLeuGln.....ProAsnLeu.... 296
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1502 TTCAGTGTGCGCTACCTGGACACCTCTTGGAAACCCCAACTTTAAGAA 1551
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
297 .....ValProGluPheArgGlyValMetLysValAspIleAsnLeuGln 311
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1552 AGTGTGCACAGGAATGAAAGCTTGAAGAAACTATGTCCAGGACAGCA 1601
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 LysValAspIleAspGlnCysSerSerAspGlyTrpPheSerGlyThrHi 328
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1602 AAATGGGATTGTCACTCAATGCCATGTATGCCATGGCAGATGGCTGCAG 1651
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 sLysCys...His.....LeuAsnAsnSerGluCysMe 338
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1652 AACATGCACCATCTCTGTCTCCGGCCATGTGGGCTGTGTGATGCTAT 1701
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
338 t...ProIleLysGly.....LeuGlyPheValLeu..... 347
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1702 GAAACCCATTGATGGCAGGAGCTCTCTGGATTCTCTCATCAATCTCTTT 1751
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
347 ..... 347
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1752 TTTCGAGAGTCTGGAGAGGAGTGTGTTCGATGAGAGGGGGATGCT 1801
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
348 ...GlyAlaTyrGluCysIle..... 353
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1802 CCCGGAAGGTATGACATTTATGAATCTGCAGTACACAGAAGCTAATCGCTA 1851
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354 .....CysLysAlaGlyPheTyrHisProGlyValLeuProValAsnA 368
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1852 TGACTATGTCCAGCTGGGACCTGGCATGAAGAGTGTCTGAATATTGATG 1901
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
368 snPheArgArgGlyProAspGlnHisIleSerGlySerThrLysAsp 384
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1902 ATTACAAATCCAG.....ATGAACAAAGCGGAATGGTACGATCT 1942
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
385 ValSerGluGlu..... 388
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1943 GTTCGAGTGAAGCTTCTTAAAGGTCAGATTAAAGTCTACGGAAGG 1992
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1993 AGAAGTGAGCTGCTGTGGATCTGCAGCGCTGCAAGAGAATGAGTTTG 2042
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
399 .....CysProPheCys..... 402
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2043 TGCAGGACGAGTTCACTGCAGAGCCTGTGACCTGGGGTGGTGGCCCAAC 2092
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
403 AlaAspAspSerProCysPheValGlnGluAspLysTyrLeuArg..... 417
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2093 CGAGAGTCAAGGCTTGAGCCCATCTCTGCTCCGTATCTTGAAGTGAG 2142
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
418 .....LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuL 430
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2143 TGACATAGATCTATCATAGCCATCGCTTTCTTGCCTGGGCATCTCTCG 2192
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
430 euAspPheValSerMetLeuValTyrHisPheArgLysAlaLysSer 446
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2193 TGACGCTGTTGTCACTCATCTTCTGTACCGGGACACACCGGTG 2242
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
447 IleArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLe 463
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2243 GTCAATCTCCAGTAGGGAGCTCTCTATATCATCTTCTGGCTGGTATTT 2292
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
463 uLeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheA 480
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2293 CCTCGGCTATGTGCCCTTTCACCTCTCATCGCAACACTACTACCAT 2342
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
480 rgCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyr 496
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

2343 CTGCTACCTCCAGCCCTCCCTAGTTGGCTCTCTCTGCCATGTGCTAC 2392
497 GlyThrValThrLeuLysLeuHisArgValLeuLysVal..... 509
2393 TCTGCTTTAGTACCAACCAACCACTCGATTGACGACATCTCTGGTGGCAG 2442
-510 .....PheLeuSerArgThrAlaG 516
2443 CAAGAAGAAGATCTGCACCGGAAAGCCCAAGATTTCATGAGCGCTTGGGCC 2492
516 LArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAlaVal 532
2493 AAGTGATCATAGCCCTCATCTGATTAGTGTACAGCTAACACTAGTGGTG 2542
533 IleLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerSerValCys 549
2543 ACCTTGATCATC..... 2554
549 sGlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerA 566
2555 .....ATGGAGCCTCCCATGCCATTTTGTCTACCGAGTATCAAGG 2597
566 sPheIleLeuPheAsnMetCysLeuIleAspArgTrpAspTyrMetThr 582
2598 AA.....GTCTACCTTATCTGCAATACCAACGACCTGGGTGATGGCC 2641
583 AlaValAla...GluPheLeuPheLeuLeuTrpGlyValTyrLeuCysty 598
2642 CTGTGGGTAGCAANTGGACTCTCTCATCATGAGCTGTACTACTATGCGCTT 2691
598 rAlaValArgThrValProSerAlaPheHisGluProArgTyrMetAlaV 615
2692 CAAGACCCGCAACGTCGCGGCCAACTTCAATGAGGCTAAATACATCGCCT 2741
615 aAlaValHisAsnGluLeuIleIleSerAlaIlePheHisThrIleArg 631
2742 TCACCATGTACACTACCTGCATCTCTGGCTGGCTTCTGCTCCCATTTAC 2791
632 PheValIleAlaSerArgLeuGlnSerAspTrpMetLeuMetLeuTyrPh 648
2792 TTT.....GGGAGCAACTACAGATCATCATCTACCTG 2823
648 eAlaHisThrHisLeuThrValThrValThrIleGlyLeuLeuIleP 665
2824 CTTGCGGTGAGCCTCAGTGTACGCGTGGCCCTGGGTGCATGTTACTC 2873
665 rOlys.....PheSerHisSerSerAsnAsnProArgAsp 677
2874 CGAAGATGTACATCATCATTTGCAACCTGAGAGGACGTCGCGAGTGC 2923
678 IleAlaThrGluAlaTyrGluAspGluLeuAspMetGly..... 690
2924 TTCAGACC.....TCTGATGTTCTCCGATGCACGTCGGTGTATGG 2964
691 .....ArgSerGlySerTyrLeuAsnSerSerIleAsnSerA 703
2965 CAACTGCGCGTCCGCTCCACACCTTCTCTCAAC..... 2998
703 IaTrpSerGluHisSerLeuAspProGluAspIleArgAspGluLeuLys 719
2998 ..... 2998
720 LysLeuTyrAlaGlnLeuGluIleTyrLysArgLysLysMetIleThrAs 736
2999 .....ATTTCGGAGAGAAAGACCGCGGCGAG 3027
736 nAsnProHisLeuGlnLysLysArgCysSer..... 746
3028 GAATGCCAATTCTAACGGCAAGTCTGTGTCTATGTTGTCGAACAGGTGAA 3077
747 ..LysLysGlyLeuGlyArgSerIleMetArgIleThrGluIlePro 762

```

```

3078 GACAGGCGCCCAAGGACAGCAGCTGTGGCAGCGCTCTCTGTGCACGTG 3127
763 GluThrValSerArgGlnCysSerLysGluAspLysGluGlyAlaAspHi 779
3128 AAGACCAAGCAGACGCGCTGTAAACAA..... 3154
779 sGlyThrAlaLysGlyThrAlaLeuIleArgLysAsnProProGluSers 796
3155 .....ACAGCGCTAATCAAA.....CCCTCACTA 3179
796 erGlyAsnThrGlyLysSerLysGluGluThrLeuLysAsnArgValPhe 812
3180 AAGATTACCAAGGCTGCGAAGAGCCTGACCTTTTCAGAT..... 3220
813 SerLeuLysLysSerHisSerThrTyrAspHisValArgAspGlnThrGl 829
3220 ..... 3220
829 uGluSerSerSerLeuProThrGluSerGlnGluGluThrThrGluA 846
3220 ..... 3220
846 snSerThrLeuGluSerLeuSerGlyLysLysLeuThrGlnLysLeuLys 862
3221 .....GCCAGCAACAGACCTTTTACAATGTGAA 3250
863 GluAspSerGluAlaGluSerThrGluSerValProLeuValCysLysSe 879
3251 GAAGAGGACAATACCCCTCTGCTCACITCAGCCCTCCAGCAGCCCTTC 3300
879 rAlaSerAlaHisAsnLeuSerSerGlyLysLysThrGlyHisProArgT 896
3301 TATGTTGTGTGACCGACGCGGCCACCCGTGGCCACACACCTCTGCTGC 3350
896 hrSerMetLeuGlnLysSerLeuSerValIleAlaSerAlaLysGluLys 912
3351 CACCCCATCTG.....ACCGAGAGAGAGACC 3376
913 ThrLeuGlyLeuAlaGlyLysThrGlnThrAlaGlyValGluGluArgTh 929
3377 CCCCTGTTCCTGGTGATTCGTCATCCCAAGGCGTTG..... 3415
929 rLysSerGlnLysProLeuProLysAspLysGluThrAsnArgAsnHis 946
3416 .....CCTCTCTCTCCGCGACGACGAG..... 3439
946 erAsnSerAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsn 962
3440 .....CCACAGCAGCGCGCC 3454
963 ProAlaGluGluProArgLysProGlnLysSerGlyIleMetLysGlnL 979
3455 CCTCAGCAGCCCGCAGCAGCCCAAG.....TCCCTGATGGACCACT 3498
979 nArgValAsnProThrThrAlaAsnSerAspLeu..... 990
3499 GCAAGGCGTAGTCCACAACTTCGGTTCGGGATTCAGATTTCCAGATTCC 3548
991 .....AsnProGlyThrThrGlnMetLysAspAsnPheAspIleGly 1004
3549 TGCTGGCAGCGCGGGGACCA.....GGA 3574
1005 GluValCysProTrpGluValTyrAspLeuThrProGlyProValProSe 1021
3575 AAC.....AGCTCGCTCTCTGTACCGCCCGCCCTCCGCC 3612
1021 rGlu.....SerLysValGlnLysHisValSerIleValAlaSerGlu 1036
3613 GCAACACCTGCAGATGCTGCCCTGCACCTGAGCAGCTTCCAGGAGGAGT 3662
1036 etGluLysAsnProThrPheSerLeuLysGluLysSerHisHisLysPro 1052
3663 CCATCTCCCTCCTGGGAGGACATCGATGATGACAGT..... 3700

```

```
1053 LysalaalaGluValCysGlnGlnSerAsnGlnLysArgIleAspLysAl 1069
      ::|||:::
3701 .....GAGAGATTCAAGCTCCT 3717
1069 aGluValCysLeuTrpGluSerGlnGly.....GlnSerIleLeuG 1083
      ::|||:::
3718 GCAGAGATTGCTGTACAGCGCCGAGGAGACACCCAGAGAGATGAATTGG 3767
1083 luAspGluLysLeuLeuSerLysThrProValLeuProGluArgAla 1099
      ::|||:::
3768 AAGAGGAGGAGGACCTGCCACGCCAGCAAGCTGACCCCTGAGGATTCT 3817
1100 LysGluGluAsnGlyGlnProArgAlaAlaAsnValCysAlaGlyG1 1116
      ::|||:::
3818 CCTGCCCTGAGCGCTCTCTCTCTCCGAGATTCCGTGGCTCTGGCAG 3867
1116 nSerGluGluLeuProPro.....LysAlaValAlaSerLysThrGluA 1131
      ::|||:::
3868 CTCAGTCCCGAGTCCCGCTGTACTGAGTCGGTCTCTGCACCCCTCAA 3917
1131 snGluAsnLeuAsnGlnIleGlyHisGlnGluLysLysThrSerSer 1147
      ::|||:::
3918 ATGTAACCTACGCTCTGCTCATCTGAGGGACTACAAAGAGCTCTTCC 3967
```

seq_name: /cqn2_6/ptodata/2/ina/6B_COMB.seq:us-08-660-148-1

seq_documentation_block:

Sequence 1, Application US/08660148

Patent No. 6211353

GENERAL INFORMATION:

APPLICANT: Burnett, J. P.

APPLICANT: Mayne, Nancy G.

APPLICANT: Sharp, Robert L.

APPLICANT: Snyder, Yvonne M.

TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND

TITLE OF INVENTION: RELATED NUCLEIC ACID COMPOUNDS

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Eli Lilly and Company

CITY: Indianapolis

STATE: Indiana

COUNTRY: United States of America

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/660,148

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/282,853

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Gaylo, Paul J.

REGISTRATION NUMBER: 36,808

REFERENCE/DOCKET NUMBER: X-9419

TELECOMMUNICATION INFORMATION:

TELEPHONE: (317) 276-0756

TELEFAX: (317) 276-3861

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4207 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 460...3999
US-08-660-148-1

alignment_scores:

Quality: 216.50 Length: 1067
Ratio: 0.454 Gaps: 42
Percent Similarity: 44.705 Percent Identity: 18.650

alignment_block:

US-09-775-181-2 x US-08-660-148-1 ..

Align-seg 1/1 to: US-08-660-148-1 from: 1 to: 4207

```
219 GluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHi 235
      ::|||:::
1456 GATGTCAAGTGGTTTGATGATTATTATCTGAAGCTCCGGCCAGAAACAAA 1505
235 sArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrpArg. 251
      ::|||:::
1506 CCACCGAAACCT.....TGGTTTCAAGAAATTTTGGCAGC 1540
252 .....ArgLysAspGlyLeuGlyAspLysSerHisPhe 263
      ::|||:::
1541 ATCGTTTTCAGTCCCGACTGGAAGGGTTTCCACAGGAGAACAAATAC 1590
264 LysTrpSerProTyrLeuGluCysGluAsnGly..... 275
      ::|||:::
1591 AACAGAGCT.....TGCATAGTCTCTGACTCTGAAAC 1625
276 .....SerTyrLysProGlyTrpLeuVal...ThrLeuSerS 287
      ::|||:::
1626 ACATCATGTTTCAGGATTCACAAATGGGATTGTGATCAACGCCATCTATT 1675
287 erAlaIleTyrGlyLeu.....GlnProAsnLeuValProGluPhe 300
      ::|||:::
1676 CGATGGCTATGGCTCCACACATGCAGATGTCCCTCTGCCAGCGCTAT 1725
301 ArgGlyVal.....MetLysValAspI1 308
      ::|||:::
1726 GCAGGACTCTGTGATGCCATGACCAATTTGATGGACGGAACCTTTTGA 1775
308 eAsnLeuGlnLysValAspIleAspGlnCysSerSerAspGlyTrpPheS 325
      ::|||:::
1776 GTCCCTGATGAAACCAATTTTACTGGGGTTTCTGGAGATACGATCCTAT 1825
325 erGlyThrHisLysCysHisLeuAsnSerGluCysMetProIleLys 341
      ::|||:::
1826 TCGATGAGAATGGAGACTCTCCAGAAAGGATGAATTAATTTTCAAG 1875
342 GlyLeuGlyPheValLeuGlyAlaTyrGluCysIleCysLysAlaGlyPh 358
      ::|||:::
1876 GAAATGGGA.....AAAGATTACTT 1895
358 eTyrHisProGlyValLeuProValAsnAsnPheArgArgGlyProA 375
      ::|||:::
1896 TGATTATATCAAGCTTGGAAGTTGGGACAATGGAGAATTAATAATGGATG 1945
375 spGlnHisIle.....SerGlySerThrLysAspValSerGlu 387
      ::|||:::
1946 ATGATGAAGTATGGTCCCAAGAAAAGCAACATCATCATCTGTGTGCAGT 1995
388 Glu..... 388
1996 GAACCATGTGAGAAAGCCAGATCAAGTGTCCGAAAGGAGNAGTCAG 2045
389 ....AlaTyrValCysLeuProCysArgGly..... 398
      ::|||:::
2046 CTGTTGTTGGACCTGTACACCTTGTGAAGGAGAAATGAGTATGCTTTTGTG 2095
399 .....CysProPheCys.....AlaAspAsp... 405
      ::|||:::
2096 AGTACACATCAAGGCATGCCAACTGGGGTCTTGGCCCACTGATGATCTC 2145
```

406	SerProCysPheValGlnGluAspLysTyrLeuArg.....	417
2146	ACAGGTGTGACATTGATCCAGTACAGTATCTTCGATGGGTGACCCCTCA	2195
418LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLeuAspPheV	433
2196	ACCATTTGACGCTGGTGTTCCTTCGCTGGCTGGCCGCCACCCCTGT	2245
433	alSerMetLeuValValTyrHisPheArgLysAlaLysSerIleArgAla	449
2246	TTGTACTGTAGTCTTCATCATTTACCGTGATACACAGTAGTCAAGTCC	2295
450	SerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeuLeuLeuTyr	466
2296	TCAAGCAGGGAACCTCTGCTACATTATTCCTTGTGTGCATCTGCTGGCTA	2345
466	rPheProValValIleLeuTyrPheGluProSerThrPheArgCysIleLe	483
2346	CTTATGTACTCTTCGCTCATTCCTGCAAGCCCAACAGATTTACTGCTACC	2395
483	euleuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrGlyThrVal	499
2396	TTCAGAGAATTGGCATTGGTCTCTCCACGCCATGAGTACTACGCCCTT	2445
500	ThrLeuLysLeuHisArgValLeuLysVal.....	509
2446	GTAACAAAGACCAACCGTATTGCAAGGATCCTGGCTGGCAGCAAGAA	2495
510PheLeuSerArgThrAlaGlnArgIleP	519
2496	GATCTGTACCAAAAGCCAGATTCATGAGTGCCTGTGCCAG.....	2538
519	roTyrMetThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLeu	535
2539CTA	2541
536	ValValPheTrpPheLeuIleGlyTrpThrSerValCysGlnAsnLe	552
2542	GTCATTGCTTTCATCTCATA.....TGCATCCAGTT	2573
552	uGluLysGlnIleSerLeu.....IleGlyGlnG	562
2574	GGGCATCATCTGTGGCCCTTTTATATGGAGCCTCCTGACATATGTCATG	2623
562	lyLysThrSerAspHisLeuIlePheAsnMetCysLeuIleAspArgTrp	578
2624	ACTACCAAGCATTCGANGTCTACCTGATCTGTAAACACCACCACCTA	2673
579	AspTyrMetThrAlaValAla...GluPheLeuPheLeuTrpGlyVa	594
2674	GGAGTGTCTACTCCACTTCGATACAAATGGATTGTTCATTTGAGCTGCAC	2723
594	lTyrLeuCysTyrAlaValArgThrValProSerAlaPheHisGluProA	611
2724	CTTCTATGGCTTCAAGACAGAAATGTTCAGCTAACTTCAACAGGGCCA	2773
611	rgTyrMetAlaValAlaValHisAsnGluLeuIleIleSerAlaIlePhe	627
2774	AGTATATGCCCTTCACATATGACAGCACTGCATATATATGCTAGCTTTT	2823
628	HisThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTrpMetLe	644
2824	GTGCCAATCTACTTT.....GGCACCAACTACAAAT	2855
644	uMetLeuTyrPheAlaHisThrHisLeuThrValThrValThrIleGlyL	661
2856	CATCACCATGTGTTTCTCGGTGAGCTTCAGTGCCACAGTGGCCCTAGGCT	2905
661	euleuLeuIleProLys.....PheSerHisSerSerAsnAsn	673
2906	GCATGTTTGTGCCGAAGGTGTACATCATCTTGGCCAAACACGAGAGAAAC	2955
674	ProArgAspPheIleAlaThrGlu...AlaTyrGluAspGlnLeuAspMe	689
2956	GTGGCAGCGCCTTACCACATCTACCGTGGTGGCATGTAGTATAGGGA	3005
689	tGlyArgSerGlySerTyrLeu.....AsnSerSerIleAsnSerAlat	704
3006	TGGCAGTCTATCTCTCCGACGCCAGCAGATCCACAGCAGCTAGTCAACCTGT	3055
704	rpSerGluHisSerLeuAspProGluAspIleArgAspGluLeuLysLys	720
3056	GGAAGAAGGGGCTCTCTGCGGGAACCTTAAGTTTCCAATGGAATAATCC	3105
721LeuTyrAl	723
3106	GTACAGTGGGCCAGAAATGAGAAGACAGCGCGGGGCGACACCTGTGGCA	3155
723	agInLeuGluIleTyrLysArgLysLysMetIleThrAsnAsnProHisL	740
3156	GCCTCTGTCATCCATCAACAAGAAAGAAAACCCCAACCAACGCGCG	3205
740	eUGlnLys.....LysArgCysSerLysLysGlyLeuGlyArgSer	753
3206	TCATCAAGCCCTTCCCCAAGACGACGAGCGCTGGCTGGCGCGTGC	3255
754	IleMetArgIleThrGluIleProGluThrValSerArgGlnCysSe	770
3255	3255
770	rLysGluAspLysGluGlyAlaAspHisGlyThrAlaLysGlyThrAlaL	787
3256GCTGGCGCAGGC...GGGAGCGCTGGGGCGGTGGGG	3289
787	euleArgLysAsnProGluSerSerGlyAsnThrGlyLysSerLys	803
3290	CCAGGGGGGTGGGGCTGGCGAGCGCGCGCCAGCGCGGCCGAGTCC	3339
804	GluGluThrLeuLysAsnArgValPheSerLeuLysLysSerHisSerTh	820
3340	CCAGAGCGCGGCCCAAGGCGCTGTAT.....	3366
820	rTyrAspHisValArgAspGlnThrGluGluSerSerSerLeuProThrG	837
3367GATGTGGCGAGGCTGAGGAGCATCTTCCGCGCGC	3400
837	luSerGlnGluGluThrThrGluAsnSerThrLeuGluSerLeuSer	853
3401	CCCGCGCGCGCTCACCGTGCCTCATCAGCAGCTGAGCCACCGCGCGG	3450
854	GlyLysLysLeuThrGlnLysLeuLysGluAspSerGluAlaGluSerTh	870
3451	GGC.....TCGGCCACGCCACGAGCAGCATGTCCCGCTCGCT	3488
870	rGluSerValProLeuValCysLysSerAlaSerAlaHisAsnLeu...Se	886
3489	GCATCGAGCGCTGTGGCGCGCAGCAGCTCTCGCAGGGCTCCCTCATGG	3538
886	rSerGluLysLysThrGlyHisProArg...ThrSerMetLeuGlnLysSer	902
3539	AGCATATCAGCATGTGGTTCACCGCTTTCACGGCCCAACATCAGCAGCTC	3588
903	LeuSerValIleAlaSerAlaLysGluLysThrLeuGlyLeuAlaGlyL	919
3589	AACTCCATGATGCTGTCCACCGCGGCCCGCCCGCGGTGGCGGCC	3638
919	ysThrGlnThrAlaGlyValGluArgThrLysSerGln.....	932
3639	GCCTCTGCTCTACCTGATCCCAAGAGATCCAGTTGCCACGACCA	3688
933	...LysProLeuProLysAspLysGluThrAsnArgAsnHisSerAsnSe	948
3689	TGACGACCTTTGCCGAATTCACCGCTCTCCCGGCCCATCGAAGTACGGG	3738
948	rAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsnProAlaG	965

```
3739 GCGCCGAGCCCGCGGAGGCGCCAGCGGCTGGGACGCGCGCCGGGA 3788
      :::::
965 luGluProArgLysProGlnLysSerGlyIleMetLysGlnArgVal 981
      ::|||
3789 GAGCCCGCGG.....CCGGTCCGAGGCGTG 3814
      ::|||
982 AsnProThrAlaAsnSerAspLeuAsnProGlyThrThrGlnMetLys 998
      ::|||
3815 CGGCG.....CCAGCCAGACC..... 3832
      ::|||
998 sAspAsnPheAspIleGlyGluValCysProTrpGluValTyrAspLeuT 1015
      ::|||
3833 .....TGGAGGAGCTGGTGGGCTCT 3851
      ::|||
1015 hrProGlyProValProSerGluSerLysValGlnLysHisValSerIle 1031
      ::|||
3852 CACCCCGCGTCCCTTCAGAGACTCGGTGGACTCGGGGAGCACACCC 3901
      ::|||
1032 ValAlaSerGluMetGluLysAsnProThrPheSerLeuLys..... 1045
      ::|||
3902 CCAACTCGCAGTGTCCGAGTCGGCCCTCTGTATCCGCTGCTCCCAA 3951
      ::|||
1046 .....GluLysSerHisLysLysProLysAlaAlaG 1056
      ::|||
3952 TATGACACTTTCATAAGAGATTACACTCAGAGCTCTCGCGCGTGTG 4001
      ::|||
1056 luValCysGlnGlnSerAsnGlnLysArgIleAspLysAlaGluValCys 1072
      ::|||
4002 AA.....TGT 4006
      ::|||
1073 LeuTrpGluSerGlnGlyGlnSerIleLeuGluAspGluLysLeuIle 1089
      ::|||
4007 CCCTGGAAGACGC..... 4021
      ::|||
1089 eSerLysThrProValLeuProGluArgAlaLysGluGluAsnGlyGly 1106
      ::|||
4022 .....CGGCTGCGCGTGGC..... 4036
      ::|||
1106 lnProArgAlaAlaAsnValCysAlaGlyGlnSerGluGluLeuProPro 1122
      ::|||
4037 .....GAGCGAGCCCGCCG 4051
      ::|||
1123 LysAlaValAlaSerLysThrGluAsnGluAsnLeuAsnGlnIleGlyHI 1139
      ::|||
4052 .....TGTTCACACACACAAT.....GGCAA 4074
      ::|||
1139 sGlnGluLysLysThrSerSerGluGluAsnValArgGlySer 1154
      ::|||
4075 GCATAGCCGCTGTGTACGCGCCAGGGGAGGTGCCAAGGGCACC 4120
      ::|||
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:us-08-660-148-3

```
seq documentation block:
; Sequence 3, Application US/08660148
; Patent No. 6211353
; GENERAL INFORMATION:
; APPLICANT: Burnett, J. P.
; APPLICANT: Mayne, Nancy G.
; APPLICANT: Sharp, Robert L.
; APPLICANT: Snyder, Yvonne M.
; TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
; TITLE OF INVENTION: RELATED NUCLEIC ACID COMPOUNDS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,148
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/282,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J. 36,808
REGISTRATION NUMBER: X-9419
REFERENCE/DOCKET NUMBER: (317) 276-0756
TELEPHONE: (317) 276-3861
TELEFAX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4207 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: mRNA
US-08-660-148-3
```

alignment_scores:
Quality: 216.50 Length: 1067
Ratio: 0.454 Gaps: 42
Percent Similarity: 44.705 Percent Identity: 18.650

alignment_block:
US-09-775-181-2 x US-08-660-148-3 ..

Align seg 1/1 to: US-08-660-148-3 from: 1 to: 4207

```
219 GluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHI 235
      :::::|||||
1456 GAUGUCAAGUGGUUGAUGAUUAUUCUGAAGCUCGCCGAGCAACAAA 1505
      ::|||
235 sArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrpArg. 251
      ::|||
1506 CCACCGAAACCCU.....UGGUUUAAGAAGAUUUUUGGCAGC 1540
      ::|||
252 .....ArgLysAspGlyLeuGlyGlyAspLysSerHisPhe 263
      ::|||
1541 AUUUUUUUCAGUCCGAGGAGGUGUUCACAGGAGACAGCAAAUAC 1590
      ::|||
264 LysTrpSerProProTyrLeuGluCysGluAsnGly..... 275
      ::|||
1591 AACAAAGACU.....UGCAUAGUUCUCUGACUCUGAAGAAAC 1625
      ::|||
276 .....SerTyrLysProGlyTrpLeuVal...ThrLeuSerS 287
      ::|||
1626 ACAUCAUGUUCAGGAUCCAAAUGGGAUUUGAUCACGCCCAUCUAU 1675
      ::|||
287 erAlaIleTyrGlyLeu.....GlnProAsnLeuValProGluPhe 300
      ::|||
1676 CGAUGGCCUAGGCGUCCACACAUAGCAGAUUCCUCCGCGCCAGGCUAU 1725
      ::|||
301 ArgGlyVal.....MetLysValAspIle 308
      ::|||
1726 GCAGGACUCUGUGAUGCAAGCAAAUUGAUGGACGGAACUUUGGA 1775
      ::|||
308 eAsnLeuGlnLysValAspIleAspGlnCysSerSerAspGlyTrpPheS 325
      ::|||
1776 GUCCUGAAGAAAACCAUUUUUACUGGGGUUUGGAGAUACGAGAUCCUAU 1825
      ::|||
325 erGlyThrHisLysCysHisLeuAsnAsnSerGluCysMetProIleLys 341
      ::|||
1826 UCGAUGAAGAGGAGACUCUCCAGGAAGUUGAUAUAUUAUUAUUAUUA 1875
      ::|||
```



```

; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 370..3912
; OTHER INFORMATION: /product= "HUMAN MGLUR5A"
; US-08-367-264-7

alignment_scores:
  Quality: 215.00      Length: 1084
  Ratio: 0.461        Gaps: 42
  Percent Similarity: 42.989  Percent Identity: 18.358

alignment_block:
US-09-775-181-2 x US-08-367-264-7
..
Align seg 1/1 to: US-08-367-264-7 from: 1 to: 4085

219  GluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHi 235
      ::::::::::::::::::::||| |||||:::
1366  GATGTCACAGTGGTTGATGATATATCTGAAGCTCCGCCAGAACAA 1415
      ::::::::::::::::::::||| |||||:::

235  sArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrpArg 252
      ::::::::::::::::::::||| |||||:::
1416  CCACCGAAACCCCT.....TGGTTTCAGAAATTTTGGCAGC 1450
      ::::::::::::::::::::||| |||||:::

252  rLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSerProPro 268
      ::::::::::::::::::::||| |||||:::
1451  ATCGT.....TTTCAGTGGCCGACTGGAA 1473
      ::::::::::::::::::::||| |||||:::

269  TyrLeuGluCysGluAsnGlySerTyr..... 277
      ::::::::::::::::::::||| |||||:::

1474  GCGTTTCCACAGAGAACAGCAATACAAAGACTTGCAATAGTCTCT 1523
      ::::::::::::::::::::||| |||||:::

278  .....LysProGlyTrpLeuVal. 283
      ::::::::::::::::::::||| |||||:::

1524  GACTCTGAAACACATCATGTTCAGGATTCCAAATGGGATTTGTGATCA 1573
      ::::::::::::::::::::||| |||||:::

284  ...ThrLeuSerSerAlaIleTyrGlyLeu.....GlnProAsnLeu 296
      ::::::::::::::::::::||| |||||:::

1574  AGCCATCATTTTCGATGGCCTATGGCTCCACACATGCAGATGTCCCTC 1623
      ::::::::::::::::::::||| |||||:::

297  ValProGluPheArgGlyVal.....Me 304
      ::::::::::::::::::::||| |||||:::

1624  TGCCCGAGGTATGCAGGACTCTGTGATGCCATGAAGCCAAATGTATGGAGC 1673
      ::::::::::::::::::::||| |||||:::

304  tLysValAspIleAsnLeuGlnLysValAspIleAspGlnCysSerSera 321
      ::::::::::::::::::::||| |||||:::

1674  GAAACTTTTGGAGTCCCTGATGAAACCAATTTTACTGGGGTTTCTGGAG 1723
      ::::::::::::::::::::||| |||||:::

321  spGlyTrpPheSerGlyThrHisLysCysHisLeuAsnAsnSerGluCys 337
      ::::::::::::::::::::||| |||||:::

1724  ATACGATCCTTCGATGAGATGAGAGACTCTCCAGGAAGGATGAAATA 1773
      ::::::::::::::::::::||| |||||:::

338  MetProIleLysGlyLeuGlyPheValLeuGlyAlaTyrGluCysIleCy 354
      ::::::::::::::::::::||| |||||:::

1774  ATGAATTTCAAGGAATGGGA..... 1794
      ::::::::::::::::::::||| |||||:::

354  sLysAlaGlyPheTyrHisProGlyValLeuProValAsnAsnPheArg 371
      ::::::::::::::::::::||| |||||:::

1795  .AAGATTACTTTGATATATCAACGTTTGAAGTTGGGACAAATGGAGAA 1843
      ::::::::::::::::::::||| |||||:::

371  rArgGlyProAspGlnHisIle.....SerGlySerThrLys 383
      ::::::::::::::::::::||| |||||:::

1844  TAAATATGGATGATGATGAAGTATGGTCCAGAAAGCAACATCATCAGA 1893
      ::::::::::::::::::::||| |||||:::

```

```

384  AspValSerGluGlu..... 388
      ::::::::::::::::::::||| |||||:::
1894  TCTGTGTGCAGTGAACCATGTGAGAAAGCCAGATCAAGGTGATCCGAAA 1943
      ::::::::::::::::::::||| |||||:::
389  .....AlaTyrValCysLeuProCysArgGluGly.... 398
      ::::::::::::::::::::||| |||||:::
1944  GGGAGAAGTCAGCTGTGTGTGGACCTGTACACCTGTGTAAGAGAATGAGT 1993
      ::::::::::::::::::::||| |||||:::
399  .....CysProPheCys..... 402
      ::::::::::::::::::::||| |||||:::
1994  ATGTCTTTGATGATGATACATGCAAGGCATGCCAACTGGGGTCTTGGCCC 2043
      ::::::::::::::::::::||| |||||:::
403  AlaAspAsp...SerProCysPheValGlnGluAspLysTyrLeuArg... 417
      ::::::::::::::::::::||| |||||:::
2044  ACTGATGATCTCACAGGTGTGACTTGATCCAGTACAGATATCTTCGATG 2093
      ::::::::::::::::::::||| |||||:::
418  .....LeuAlaIleSerPheGlnGlyLeuCysMetL 429
      ::::::::::::::::::::||| |||||:::
2094  GGGTGACCTGAACCAATGTCAGCTGTGTGTTCCTGCTGCCCTGGCTCC 2143
      ::::::::::::::::::::||| |||||:::
429  euLeuAspPheValSerMetLeuValValTyrHisPheArgLysAlaLys 445
      ::::::::::::::::::::||| |||||:::
2144  TGGCCACCTGTTTGTGTACTGTAGTCTTCATCATTTACCGTGATACACCA 2193
      ::::::::::::::::::::||| |||||:::
446  SerIleArgAlaSerGlyLeuIleLeuGluThrIleLeuPheGlySe 462
      ::::::::::::::::::::||| |||||:::
2194  GTAGTCAAGTCTCAAGCAGGAACTCTGCTACATATCTTGTCTGGCAT 2243
      ::::::::::::::::::::||| |||||:::
462  rLeuLeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrP 479
      ::::::::::::::::::::||| |||||:::
2244  CTGCTGGGCTACTTATGTATGCTTTCCTGCTCATTTGCGAAGCCCAACAGA 2293
      ::::::::::::::::::::||| |||||:::
479  heArgCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrVal 495
      ::::::::::::::::::::||| |||||:::
2294  TTTACTGTACTCTCAGAAATGGCATTTGGTCTCTCCCGACCCATGAGC 2343
      ::::::::::::::::::::||| |||||:::
496  TyrGlyThrValThrLeuLysHisArgValLeuLysVal..... 509
      ::::::::::::::::::::||| |||||:::
2344  TACTCAGCCCTTGTAACAAGACCAACCGTATTCGCAAGGATCCTGGCTGG 2393
      ::::::::::::::::::::||| |||||:::
510  .....PheLeuSerArgThr 515
      ::::::::::::::::::::||| |||||:::
2394  CAGCAAGAAGAAGATCTGTACCCCAAGCCAGATTCATGATGCTGTG 2443
      ::::::::::::::::::::||| |||||:::
515  LglnArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAla 531
      ::::::::::::::::::::||| |||||:::
2444  CCCAG..... 2448
      ::::::::::::::::::::||| |||||:::
532  ValIleLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerSerVa 548
      ::::::::::::::::::::||| |||||:::
2449  .....CTAGTGATTTGCTTTCATTTCTCAT..... 2472
      ::::::::::::::::::::||| |||||:::
548  lCysGlnAsnLeuGluLysGlnIleSerLeu..... 558
      ::::::::::::::::::::||| |||||:::
2473  .TGATCCAGTTGGGCATCATCGTTGCCCTCTTTATAATGGAGCCTCTG 2521
      ::::::::::::::::::::||| |||||:::
559  ...IleGlyGlnClyLysThrSerAspHisLeuIlePheAsnMetCysLeu 574
      ::::::::::::::::::::||| |||||:::
2522  ACATAATGCATGACTACCCCAAGCATTCGAGAAGTCTACCTGATCTGTAA 2571
      ::::::::::::::::::::||| |||||:::
575  lLeAspArgTrpAspTyrMetThrAlaValAla...GluPheLeuPheLe 590
      ::::::::::::::::::::||| |||||:::
2572  ACCACCAACTAGGAGTTGTCACTCCACTTGGAAACAAATGGATTTGTGAT 2621
      ::::::::::::::::::::||| |||||:::
590  uLeuTrpGlyValTyrLeuCysTyrAlaValArgThrValProSerAlaP 607
      ::::::::::::::::::::||| |||||:::
2622  TTTGAGTGCACCTTCTATGCTTCAAGCAGAAATGTTCCAGCTAACT 2671
      ::::::::::::::::::::||| |||||:::
607  heHisGluProArgTyrMetAlaValAlaValHisAsnGluLeuIleIle 623
      ::::::::::::::::::::||| |||||:::
2672  TCCCGGAGCCCAAGTATATCCCTTCACATGTACACGACCTGCATTATA 2721
      ::::::::::::::::::::||| |||||:::

```

624 SerAlaIlePheHisThrIleArgPheValLeuAlaSerArgLeuGlnSe 640
2722 TGGCTAGCTTTTGTTCCTCACTACTTT.....GGCAG 2753
640 rAsPTrpMetLeuMetLeuThrPheAlaHisThrHisLeuThrValThrV 657
2754 CAACACAAATCATCATCAGCTGTCTTTCGTGCTCAGCTCAGTCCACAG 2803
657 alThrIleGlyLeuLeuLeuLeuLeuProLys.....PheSerHis 669
2804 TGGCCCTAGGCTGCTGCTTTGTGCGAAGGTGTATCATCTCTGCGCAAA 2853
670 SerSerAsnProArgAspAlaThrGlu.....AlaTyrgluAs 685
2854 CCAGAGAGAACGTCGCGAGCGCTTCCACCATCTACCGTGGTGGCGAT 2903
685 pGluLeuAspMetGlyArgSerGlySerTyrlLeu.....AsnSerSerI 700
2904 GCATGTAGGGATGCAAGTATCTCTCGCAGCCAGCAGATCCAGCAGCC 2953
700 leAsnSerAlaTrpSerGluHisSerLeuAspProGluAspIleArgAsp 716
2954 TAGTCAACCTGTGAAGAGAGAGGGCTCTCTGCGGAAACCTTAAGTTCC 3003
717 GluLeuLysLys..... 720
3004 AATGGAAATCCGTCACGTGGCCGACAGTGAAGAGCAGCCGGGGGCA 3053
721 ...LeuTyrgluAlaGlnLeuGluIleTyrgluLysLysMetIleThrA 736
3054 GCACCTGTGGCAGCGCTGTCTCATCCACATCAACAGAAAGAAACCCCA 3103
736 snAsnProHisLeuGlnLys.....LysArgCysSerLysLysGly 749
3104 ACCAAACGGCGCTCATCAAGCCCTTCCCAAGACGACGAGAGCGGTGC 3153
750 LeuGlyArgSerIleMetArgIleThrGluIleProGluThrValSe 766
3154 CTGGCGCTGGC..... 3165
766 rArgGlnCysSerLysGluAspLysGluGlyAlaAspHisGlyThrAlaL 783
3166GCTGGCGCAGGC...GGGAGCGTG 3187
783 ysGlyThrAlaLeuIleArgLysAsnProProGluSerSerGlyAsnThr 799
3188 GGGCGTGGGGCCACGCGCGTGGGCTCGCAGCGCGCGCGCCACAGC 3237
800 GlyLysSerLysGluGluThrLeuLysAsnArgValPheSerLeuLysLy 816
3238 GGGCCGAGTCCCGCAGCGCGCGCGCCCAAGCGCTGTAT..... 3276
816 sSerHisSerThrTyrgluHisValArgAspGlnThrGluGluSerSers 833
3277GATGTGGCGCAGGCTGAGGAGC 3298
833 erLeuProThrGluSerGlnGluGluThrThrGluAsnSerThrLeu 849
3299 ACTTCCCGCGCGCGCGCGCGCTCACCCTGCGCCATCAGCAGCGCTG 3348
850 GluSerLeuSerGlyLysLeuThrGlnLysLeuLysGluAspSerG1 866
3349 AGCCACCGCGCGCG.....TCGGCCACCGCCAGCAGCAGCA 3386
866 uAlaGluSerThrGluSerValProLeuValCysLysSerAlaSerAlaH 883
3387 TGTGCGCTGCTGCACTCGGAGCTGTGGCGCGCAGCAGCTCTCTCG.... 3432
883 isAsnLeuSerSerGluLysLysThrGlyHisProArgThrSerMetLeu 899
3433CAGGGCTCCCTCATG 3447
900 GlnLysSerLeuSerValIleAlaSerAlaLysGluLysThrLeuGlyLe 916

3448 GAGCAGATCAGCAGTGTGGTACCCGCTTC..... 3477
916 uAlaGlyLysThrGlnThrAlaGlyValGluGluArgThrLysSerGlnL 933
3478ACGGCAACATCAGCGAG..... 3495
933 ysProLeuProLysAspLysGluThrAsnArgAsnHisSerAsnSerAsp 949
3496CTCAACTCCATG 3507
950 AsnThrGluThrLysAspProAlaProGlnAsnSerAsnProAlaGluG 966
3508 ATGTGTCCACCGCGCGCGCGCGCTCGCGCGCGCGCTCTCTGCTC 3557
966 luProArg...LysProGlnLysSerGly..... 974
3558 GTCCTACTGATCCCAAGAGATCCAGTTGCCACGACCATGAGACCT 3607
975 IleMetLysGlnGlnArgValAsnProThrThrAlaAsnSerAspLeuAs 991
3608 TTGCGGAAATCCAGCCTCTGCGCGCATCGAAGTCACGGCGCGCTCAG 3657
991 nPro.....GlyThrThrGlnMetLysAsnPro 1001
3658 CCCGCGCAGGGCGCAGCGGTGGGGACGCGCGCGGGAGAGCCCGCG 3707
1001 heAspIleGlyGluValCysPro.....TrpGluValTyrlAsp 1013
3708 GCGCGGTCCCGAGGCTCGCGCGCGCAAGCAGACCTGGAGGAGCTGGTGG 3757
1014 LeuThrProGlyProValProSerGluSerLysValGlnLysHisValSe 1030
3758 CTCTCACCCCGCGCTCCCGCTTCAGAGACTCGGTGGACTCGGGGAGCACA 3807
1030 rIleValAlaSerGluMetGluLysAsnProThrPheSerLeuLys.... 1045
3808 ACCCCCACTCGCCAGTCTCGAGTGGCCCTCTGTATCCGCTGCTTCC 3857
1046GluLysSerHisLysProLysAla 1054
3858 CAATATACACTCTTATCATAGAGATTACACTCAGAGCTCTCTCGTGT 3907
1055 AlaGluValCysGlnGlnSerAsnGlnLysArgIleAspLysAlaGluVa 1071
3908 TGTGAA..... 3913
1071 lCysLeuTrpGluSerGlnGlyGlnSerIleLeuGluAspGluLysLeuL 1088
3914 .TGTCCCTGGAAGCAGCG..... 3931
1088 euIleSerLysThrProValLeuProGluArgAlaLysGluGluAsnGly 1104
3932CGGCTCGCGCTGCG..... 3946
1105 GlyGlnProArgAlaAlaAsnValCysAlaGlyGlnSerGluGluLeuPr 1121
3947GAGCGAGGCC 3957
1121 oproLysAlaValAlaSerLysThrGluAsnGluAsnLeuAsnGlnIleg 1138
3958 CCGG.....TGTTCACACACACACAT.....G 3980
1138 lHisGlnGluLysLysThrSerSerSerGluGluAsnValArgGlySer 1154
3981 GCAAGCATAGTCGCTGTTACGCCCGCGCGGAGAGATGCCAAGGACCC 4030
seq_name: /cgn2_6/ptodata/2/ina/6b_comb.seq:US-08-660-148-4

seq_documentation_block:

; Sequence 4, Application US/08660148

; Patent No. 6211353

; GENERAL INFORMATION:

2496 GATCTGTACCAAAAGCCAGATTCATGAGTGCCTGTGCCAG..... 2538
519 roTyrMetThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLeu 535
2539CTA 2541
536 ValValPheThrPheLeuIleGlyTyrThrSerSerValCysGlnAsnLe 552
2542 GTGATTGCTTTCATCTCATA.....TGCAATCCAGTT 2573
552 uGluLysGlnIleSerLeu.....IleGlyGlnG 562
2574 GGGCATCATGCTGCCCTCTTTATAATGAGGCTCTGACATAATGCATG 2623
562 lYsThrSerAspHisPheAsnMetCysLeuIleAspArgTyr 578
2624 ACTACCAAGATTCGAGAGATTCACCTGATCTGTAACACACCACTA 2673
579 AspTyrMetThrAlaValAla...GluPheLeuPheLeuTyrGlyVa 594
2674 GGAGTCTGCTCACTCCACTTGGATCAATGGATGTTGATTTGAGCTGCAC 2723
594 lTyrLeuCysTyrAlaValArgThrValProSerAlaPheHisGluProA 611
2724 CTTCTATCGGTTCAAGACCAAGATGTTCCAGCTAACTTCAACGAGGCCA 2773
611 rgTyrMetAlaValAlaValHisAsnGluLeuIleIleSerAlaIlePhe 627
2774 AGTATATCGGCTTCACAAATGTACAGACCTGCAATATATATGCTAGCTTT 2823
628 HisThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTyrMetLe 644
2824 GTGCCATCTACTTT.....GGCAGCAACTACAAAT 2855
644 uMetLeuTyrPheAlaHisThrHisLeuThrValThrValThrIleGlyL 661
2856 CATCACCATGTGTTTCTCGGTACGCTCAGTCCACAGTGGCCCTAGGCT 2905
661 euLeuLeuIleProLys.....PheSerHisSerSerAsnAsn 673
2906 GCATGTTTGTGCGAAGGTGTACATCATCTCTGCGCAACACAGAGAGAAC 2955
674 ProArgAspAlaIleThrGlu...AlaTyrGluAspGluLeuAspMe 689
2956 GTGCGAGCGCTTCACCACTATCCGTTGCGGATGCTAGTGGGA 3005
689 tGlyArgSerGlySerTyrLeu.....AsnSerSerIleAsnSerAlat 704
3006 TGGCAAGTCATCTCCGCGAGCCAGCAGATCCAGCAGCTAGTCAACCTGT 3055
704 rpSerGluHisSerLeuAspProGluAspIleArgAspGluLeuLysLys 720
3056 GGAAGAAAGGGGCTCTCTGGGAAACCTTAAAGGTACAAGACAGAGA 3105
721 Leu.....TyrAlaGlnLeuIleTyrLysArgLysLysMetIl 734
3106 CTGCCCCAGACAGACTCGGAATAGAGTGTTCACCCCAAGGAGATAT 3155
734 e.....ThrAsnAsnProHisLeuG 741
3156 GGGGAATGGTGGGAGACAACATGACAGTTCCAATGGAAATCCCTCA 3205
741 lnLysLysArgCysSerLysLysGlyLeuGlyArgSerIleMetArgArg 757
3206 CGTGGGCCCAAGATGGAAGACGACCGGGGCGAGCACCTGTGCGACGCG 3255
758 IleThr.....GluIlePro..... 762
3256 CTGTCCATCCACATCACAGAAAGAACCCCAACCAACGCGCGCTAT 3305
763GluThrValSerArgGlnCysSerLysGluAspL 774
3306 CAAGCCCTTCCCCAAGACGACGAGACCGGTGGCGCTGGCGCTGGCGCTG 3355

774 ysGluGlyAlaAspHisGly.....ThrAlaLysGlyThr 785
3356 GCGCAGCGGGAGCGGTGGGGCGTGGGCCACGGCGGTGGCGGTGC 3405
786 AlaLeuIleArgLysAsnProGluSerSerGlyAsnThrGlyLysSe 802
3406 GCAGCGCCCGCCAGCGCGCGCGAGTCCACAGCCCGCCCAAGGC 3455
802 rLysGluGluThrLeuLysAsnArgValPheSerLeuLysLysSerHis 819
3456 G..... 3456
819 erThrTyrAspHisValArgAspClnThrGluGluSerSerLeuPro 835
3457CTGTATGATGTGGCGAGGCTGAGGAGCACCTTCCCG 3492
836 ThrGluSerGlnGluGluThrThrGluAsnSerThrLeuGluSerLe 852
3493 GCGCCCGCGCGCGCTCACCGTCCGCCATCAGCACGCTGAGCCACCG 3542
852 uSerGlyLysLysLeuThrGlnLysLeuLysGluAspSerGluAlaGlu 869
3543 CGCGGGC.....TCGGCCAGCGCAGCGAGCATGTGCCGT 3580
869 erThrGluSerValProLeuValCysLysSerAlaSerAlaHisAsnLeu 885
3581 CGGTGCACTCGGAGCTGTGGCGCGCAGAGCTCTCGCAGGGCTCCCTC 3630
886 .SerSerGluLysLysThrGlyHisProArg..ThrSerMetLeuGlnL 901
3631 ATGAGGAGATCAGCAGTGTGGTACCCCGCTTACGGCCCAACATCAGCA 3680
901 sSerLeuSerValIleAlaSerAlaLysGluLysThrLeuGlyLeu Ala 917
3681 GCTCAACTCCTATGATGTGTCCACCGCGGCCCGCCGCGGTGGCG 3730
918 GlyLysThrGlnThrAlaGlyValGluGluArgThrLysSerGln..... 932
3731 CCGCGCTCTGCTGCTACTGATGATCCCAAGAGATCCAGTTGCCACG 3780
933LysProLeuProLysAspLysGluThrAsnArgAsnHisSerA 947
3781 ACATGACGACCTTTCGCGAAATCCAGCCTTCCCGGCCCATCGAAGTCAC 3830
947 snSerAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsnPro 963
3831 GGGCGCGCGCAGCGCGCGCGGCGCAGCGGCTGGGAGCGCGCC 3880
964 AlaGluLupProArgLysProGlnLysSerGlyIleMetLysGlnGlnAr 980
3881 GGGAGAGCCCGCGG.....CCGCTCCCGAG 3906
980 gValAsnProThrThrAlaAsnSerAspLeuAsnProGlyThrThrGlnM 997
3907 GCTGCGCGCG.....CCAAGCCAGACC..... 3928
997 etLysAspAsnPheAspIleGlyValCysProTyrGluValTyrAsp 1013
3929TGGAGGAGCTGCTGG 3943
1014 LeuThrProGlyProValProSerGluSerLysValGlnLysHisValSe 1030
3944 CTCTCACCCCGCGCTCCCTTTCAGAGACTCGGTGGACTCGGGAGACACA 3993
1030 rIleValAlaSerGluMetGluLysAsnProThrPheSerLeuLys..... 1045
3994 ACCCCCAACTCGCAGTGTGGAGTGGCGCTCTGTATCCGCTGCTCC 4043
1046GluLysSerHisLysLysProLysAla 1054
4044 CAATATGACACTCTTATATAGAGATTTACACTCAGAGCTCTCTCGCGC 4093

1055 AlaGluValCysGlnGlnSerAsnGlnLysArgIleAspLysLysAlaGluVa 1071
:::||||
4094 TGTGA..... 4099
1071 lCysLeuTrpGluSerGlnGlnSerIleLeuGluAspGluLysLeuL 1088
||| |||:::||||
4100 .TGTCCCTGGAAAGCAGC..... 4117
1088 eulleSerLysThrProValLeuProGluArgAlaLysGluGluAsnGly 1104
||| |||
4118CGGCTGCGCGTGG..... 4132
1105 GlyGlnProArgAlaAlaAsnValCysAlaGlyGlnSerGluGluLeuPr 1121
|||::: ||
4133GAGCGGAGCCC 4143
1121 oProLysAlaValAlaSerLysThrGluAsnGluAsnLeuAsnGlnIleG 1138
||| |||:::||||
4144 CCG.....TGTTACACACACAAAT.....G 4166
1138 lYHisGlnGluLysLysThrSerSerGluGluAsnValArgGlySer 1154
|| ::::: |||:::||||
4167 GCAAGCATAGCCGCTGTTACGCCGCCAGGGGAAGTCCCAAGGCGACC 4216

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-660-148-6

seq_documentation_block:
; Sequence 6, Application US/08660148
; Patent No. 6211353
; GENERAL INFORMATION:
; APPLICANT: Burnett, J. P.
; APPLICANT: Mayne, Nancy G.
; APPLICANT: Sharp, Robert L.
; APPLICANT: Snyder, Yvonne M.
; TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
; TITLE OF INVENTION: RELATED NUCLEIC ACID COMPOUNDS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/660,148
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/282,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REFERENCE/DOCKET NUMBER: X-9419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
US-08-660-148-6

alignment_scores:
Quality: 209.50 Length: 1085
Ratio: 0.436 Gaps: 43
Percent Similarity: 44.332 Percent Identity: 18.525
alignment_block:
US-09-775-181-2 x US-08-660-148-6 ..
Align seg 1/1 to: US-08-660-148-6 from: 1 to: 4303
219 GluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuH1 235
||| |||:::||||
1456 GAUGUCAAGUGGUUGAUGAUUAUUGAAGCUCGCGCCAGAACAA 1505
235 sArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrpArg. 251
:::|||||::: |||::: |||:::
1506 CCACCGAAACCCU.....UGGUUCAAGAAUUUUGGCACC 1540
252ArgLysAspGlyLeuGlyGlyAspLysSerHisPhe 263
||| :::||||::: |||::: |||:::
1541 AUGGUUUUCAGUGCGCAGUCUGGAAGGUUUCACAGGAGAACAGAAUAC 1590
264 LysTrpSerProTrpLysLeuGluCysGluAsnGly..... 275
::: |||::: |||::: |||:::
1591 AAACAAGACU.....UGCAUAGUUCUCUGACUCUGAAAC 1625
276SerTyrLysProGlyTrpLeuVal...ThrLeuSers 287
||| ||| |||::: |||::: |||:::
1626 ACAUCAUGUUCAGGAUCCAAUUGGGAUUUGAUGAACGCCAUCUAU 1675
287 erAlaIleTyrGlyLeu.....GlnProAsnLeuValProGluPhe 300
|| ||| ||| |||::: |||::: |||:::
1676 CGAUGGCCUUAUGGGCUCCACAAUGCAGAUCCUCUGCCCGCCGCUAU 1725
301 ArgGlyVal.....MetLysValAspI1 308
|||::: |||::: |||:::
1726 GCAGGACUCUGUGAUGCCCAUUGAUGGACGGAACUUUGGA 1775
308 eAsnLeuGlnLysValAspIleAspGlnCysSerSerAspGlyTrpPheS 325
::: |||::: |||::: |||::: |||:::
1776 GUCCUGAUGAAACCAUUUACUGGGGUUUUGGAGAUAGCAUCCUAU 1825
325 ergLysHisLysCysHisLeuAsnSerGluCysMetProIleLys 341
::: |||::: |||::: |||::: |||:::
1826 UCGAUGAAGAUUGGAGACUCUCAGGAGGAUAAUAAUUAUUAUUAAG 1875
342 GlyLeuGlyPheValLeuGlyAlaTyrGluCysIleCysLysAlaGlyP 358
::: |||::: |||::: |||::: |||:::
1876 GAAUGGGA.....AAAGAUAUACUU 1895
358 eTyrHisProGlyValLeuProValAsnAsnPheArgArgGlyProA 375
|||::: |||::: |||::: |||::: |||:::
1895 UGAUUAUACAACGUUGGAGUUGGAGACAUUGGAGAAUUAUUAUUAUG 1945
375 spGlnHisIle.....SerGlySerThrLysAspValSerGlu 387
|||::: |||::: |||::: |||::: |||:::
1946 AUGAUGAAGAUUGGUCCAAGAAACCAUCAUCAUGAUCUGUGGUGCAGU 1995
388 Glu..... 388
1996 GAACAUGUGAAGAGCCAGCAUGAUGGAGUUGGAGAAAGGAGAGUUGCAG 2045
389AlaTyrValCysLeuProCysArgGluGly..... 398
::: |||::: |||::: |||::: |||:::
2046 CUGUUGUUGACCUGACACCCUUGUAAGCAGAUAGUAGUUCUUGAUG 2095
399CysProPheCys.....AlaAspAsp... 405
||| ||| |||::: |||:::
2096 AGUACACAUAGCAAGCAUGCCCAACUGGGGUCUUGGCCACACUGAUGUC 2145
406 SerProCysPheValGlnGluAspLysTyrIleuArg..... 417
::: |||::: |||::: |||::: |||:::
2146 ACAGUUGUGACUAGUCCAGAUACUUCUGAUGGGGUGAGCCUUGA 2195

418 LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLeuAspPheV 433
2196 ACCCAUUGCAGCUGUGUGUUUGCCUGCUUGCGGCCUGGGCACCCUGU 2245
433 aiSerMetLeuValValTyrHisPheArgLysAlaLysSerIleArgAla 449
2246 UUGUACUGAGUCUUAUCAUUAUACCGUGAUCACACACAGUAGUCAGUCC 2295
450 SerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeuLeuLeuTy 466
2296 UCAACGAGGGAACUCUGCAUAUAUUGCUGCGCAUCUGCGGCAUCUGCGGCU 2345
466 rPheProValValIleLeuTyPheGluProSerThrPheArCySileL 483
2346 CUUAUGUACCUCUGCCUAUUGGAAAGCCCAAACAUAUAUUGCUAUCC 2395
483 euLeuLeuTrpAlaArgLeuLeuGlyPheAlaThrValTyrGlyThrVal 499
2396 UUCAGAGAUAUGGAUUGGUCUCUCCCCAGCCAUGACUACUACAGCCUU 2445
500 ThrLeuLysLeuHisArgValLeuLysVal..... 509
2446 GUACAAAAGACCAACCGUAUUGCAAGAAUGCUGGCGACGACGAAGA 2495
510 PheLeuSerArgThrAlaGlnArgIleP 519
2496 GAUCUGUACCAAAAGCCCAGAUCAUGAGUGCGUGGCCAG..... 2538
519 roTYrMetThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLeu 535
2539 CUA 2541
536 ValValPheTrpPheLeuIleGlyTrpThrSerSerValCysGlnAsnLe 552
2542 GUGAUUGGUUUAUUCUAUA.....UGCAUCCAGU 2573
552 uGluLysGlnIleSerLeu.....IleGlyGlnG 562
2574 GGCGAUAUCUGGCCCCUUUAUUAUGGAGCCUCCUGCAUAUAUGCAUG 2623
562 lYlYSThrSerAspHisLeuIlePheAsnMetCysLeuIleAspArgTrp 578
2624 ACUACCCAAGAUUGCAGAAGUCUACCUAGUAGUAACACCAACCACCUA 2673
579 AspTYrMetThrAlaValAla...GluPheLeuPheLeuLeuTrpGlyVa 594
2674 GGAGUUGUCACUCCAGUGAUAACAAUGAUGAUGUUAUUGAGUCUGCAC 2723
594 tYrLeuCystYrAlaValArgThrValProSerAlapHeHisGluProA 611
2724 CUUCUAUGGUCAAGACCAGAAUUGUCCAGCUAACUCCAGCAGGCCA 2773
611 rgTYrMetAlaValAlaValHisAsnGluLeuIleIleSerAlaIlePhe 627
2774 AGUAUAUUGCCUUCACAAUGUACAGCACCUCAUAUAUAGGCUAGCUUU 2823
628 HisThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTrpMetLe 644
2824 GUGCCAAUUCUAUUU.....GGCAGCAACUACAAAAA 2855
644 uMetLeuTYrPheAlaHisThrHisLeuThrValThrValThIleGlyL 661
2856 CAUCACCAUGUUUUCUGGCACCUCAUGGCCAGUUGGCCUAGGCU 2905
661 euLeuLeuIleProLys.....PheSerHisSerSerAsnAsn 673
2906 GCAUGUUUGCCGAGGUGUAUCAUCAUCCUGGCCAAACACAGAGAGAAC 2955
674 ProArgAspAPileAlaThrGlu...AlaTYrGluAspGluLeuAspMe 689
2956 GUGCGACGCGCUUACACCAUACUACUGGUGGCGAUGCAUGUAGGGA 3005

689	tGlyArSerGlySerTyrLeu.....AsnSerSerIleAsnSerAlaT	704
690	: :	
3006	UGGAAGTCAUCCUCCGACGACGACAGCAUCCAGCAUCCAGUACAACCGUG	3055
704	rPserGluHisSerLeuAspProGluAspIleArgaspGluLeuLysLys	720
705	: : : : : : : : : :	
3056	GGAGAAGAGGGCUCUCUGGGGAAACCUUAGGUACAAAGACAGAGGA	3105
721	Leu.....TyrAlaGlnLeuGluIleTyrLysArgLysLysMetI	734
722	: : : : : : : : : :	
3106	CUGGCCAGCACAGUCGGAUAUAGAGUUGUACCCCAAGGGAGUAU	3155
734	e.....ThrAsnAsnProHisLeuG	741
735	: : : : : : : : : :	
3156	GGGAAUGUGGGAGAGCAACAUAAGCAGAUCCAAUGGAAAAUCCGUA	3205
741	InLysLysArgCysSerLysLysLysLysLeuGlyArgSerIleMetArg	757
742	: : : : : : : : : :	
3206	CGUGGGCCAGAUAGAAGAGCAGCGCGGGGAGCAGCACCUUGGGCAGCG	3255
758	IleThr.....GluIlePro.....	762
759	: : : : : : : : : :	
3256	CUGUCAUCCACAUACAAGAAGAAACCCCAACCAACGCGCCGUAU	3305
763GluThrValSerArgGlnCysSerLysLysGluAspL	774
764	: : : : : : : : : :	
3306	CAAGCCCUUCCCCAAGAGCAGGAGCCGUGGCCUGGCGGCGUGGCGUG	3355
774	ysGluGlyAlaAspHisGly.....ThrAlaLysGlyThr	785
775	: : : : : : : : : :	
3356	GCGCAGCGGGAGCGUGGGGGCGUGGGGCCACGCGGCGUGGCGCGUGC	3405
786	AlaLeuIleArgLysAsnProGluSerSerGlyAsnThrGlyLysSe	802
787	: : : : : : : : : :	
3406	GCAGGCGCGGCCCGCGGGCGCGAGUCCCAAGACCGCGGCCCAAGGC	3455
802	rLysGluGluThrLeuLysAsnArgValPheSerLeuLysLysSerHis	819
3456	G.....	3466
819	erThrTyrAspHisValArgAspGlnThrGluGluSerSerSerLeuPro	835
820	: : : : : : : : : :	
3457CUGUAUGAUGUGGCGGAGCAGCAGGAGCAUUCUCCG	3492
836	ThrGluSerGlnGluGluThrThrGluAsnSerThrLeuGluSerLe	852
837	: : : : : : : : : :	
3493	CGCGCCGCGCGCGCGCUCACCGUCGCCCAUCAGCAGCGUGAGCACCG	3542
852	uSerGlyLysLysLeuThrGlnLysLeuLysGluAspSerSerGluAlaGlu	869
853	: : : : : : : : : :	
3543	CGCGGGC.....UCGGCCAGCCGACGAGCAGCAUGUGGCCG	3580
869	erThrGluSerValProLeuValCysLysSerAlaSerAlaHisAsnLeu	885
870	: : : : : : : : : :	
3581	CGCUCACUGGAGCCUGUGGGCGGCGAGCAGUCCUCCAGGCGUCCUC	3630
886	..SerSerGluLysLysThrGlyHisProArg..ThrSerMetLeuGlnL	901
887	: : : : : : : : : :	
3631	AUGGAGCAGAUCAAGUGUGGUACACCGGUCCACCGCCCAACAUACAGCA	3680
901	sSerLeuSerValIleAlaSerAlaLysGluLysThrLeuGlyLeuAla	917
902	: : : : : : : : : :	
3681	GCUAACUUGAUGAUGUGUCCACCGCGGGCCCCAGCCCGCGGCGGCG	3730
918	GlyLysThrGlnThrAlaGlyValGluGluArgThrLysSerGln.....	932
919	: : : : : : : : : :	
3731	CCCCGCUUGCUGUCCUACCUAGUACCCCAAGAGAUCCAGUUGCCCAAG	3780
933LysProLeuProLysAspLysGluThrAsnArgAsnHisSerA	947
934	: : : : : : : : : :	
3781	ACCAUGAGCAGCUUUGCCGAAUCCAGCCUUGCGCGGCCCAUCCAGAGUC	3830
947	snSerAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsnPro	963

```

3831 GGGCGGCGCGACCGCGCGAGGGCGGCGAGGGCGGCGGCGCGCGCGCC 3880
964 AlaGluGluProGluProGlnLysSerGlyIleMetLysGlnGlnAr 980
3881 GGGAGAGCCCGCGG.....CCGUGCCGAG 3906
980 gValAsnProThrThrAlaAsnSerAspLeuAsnProGlyThrThrGlnM 997
3907 GUGCGCGCG.....CCAGCCAGACC.... 3928
997 etLysAspAsnPheAspIleGlyGluValCysProTrpGluValTyrAsp 1013
3929 .....UGGAGGAGCGUGGUGG 3943
1014 LeuThrProGlyProValProSerGluSerLysValGlnLysHisValSe 1030
3944 CUCUCACCGCGCGCCCUUCAGAGACUGGUGGACUGGGGAGCACA 3993
1030 rIleValAlaSerGluMetGluLysAsnProThrPheSerLeuLys.... 1045
3994 ACCCCCAACUGCGCAGUGCCGAGUGCGCCUCUGUAUCCGUGUCUCC 4043
1046 .....GluLysSerHisLysProLysAla 1054
4044 CAAUUAUGACACUCUUAUAGAGAUUACACAGAGCUCUGCCGCGC 4093
1055 AlaGluValCysGlnGlnSerAsnGlnLysArgIleAspLysAlaGluVa 1071
4094 UGUGAA..... 4099
1071 lCysLeuTrpGluSerGlnGlnSerIleLeuGluAspGluLysLeuL 1088
4100 .UGUCCUGGAAAGCAGCG..... 4117
1088 euIleSerLysThrProValLeuProGluArgAlaLysGluGluAsnGly 1104
4118 .....CGGCGUGCGGUGCG..... 4132
1105 GlyGlnProArgAlaAlaAsnValCysAlaGlyGlnSerGluGluLeuPr 1121
4133 .....GAGCGGAGCC 4143
1121 oProLysAlaValAlaSerLysThrGluAsnGluAsnGlnIleG 1138
4144 CCGG.....UGUUCACACACACAAU.....G 4166
1138 lHisGlnGluLysLysThrSerSerSerGluGluAsnValArgGlySer 1154
4167 GCAAGCAUAGCCGCGUGUUGCGGCCCGAGGGGAGGUGGCCAAGGCGACC 4216

```

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-072-574-7

```

seq_documentation_block:
; Sequence 7, Application US/08072574
; Patent No. 5521297
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 00719
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,574
; FILING DATE: 19930604
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9383
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 370..3912
; OTHER INFORMATION: /product= "HUMAN MGLUR5A"
; US-08-072-574-7

alignment_scores:
Quality: 208.00 Length: 1084
Ratio: 0.447 Gaps: 42
Percent Similarity: 42.897 Percent Identity: 18.266

alignment_block:
US-09-775-181-2 x US-08-072-574-7 ..
Align seg 1/1 to: US-08-072-574-7 from: 1 to: 4085

219 GluThrGluTrpPheHisGlyLeuArgLysTrpArgProHisLeuHi 235
1366 GATGTCAGTGGTTGATGATTATTATCTGAAGCTCCGCCAGAACAAA 1415
235 sArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrpArgA 252
1416 CCACCGAAACCCCT.....TGGTTTCAAGAAATTTTGCAGC 1450
252 tqLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSerProPro 268
1451 ATCGT.....TTTCAGTGGCGACTGGAA 1473
269 TyrLeuGluCysGluAsnGlySerTyr..... 277
1474 CGGTTTCCACAGGAGACAGCAATACAAACAGACTTGCATAGTTCTCT 1523
278 .....LysProGlyTrpLeuVal. 283
1524 GACTCTGAAAACACATCATGTTTCAGGATTCCAAAATGGGATTGTGATCA 1573
284 ...ThrLeuSerSerAlaIleTyrGlyLeu.....GlnProAsnLeu 296
1574 ACGCTATTATTCGATGGCTTGGCTTGGCTCCACACATGCAGATGTCCTC 1623
297 ValProGluPheArgGlyVal.....Me 304
1624 TGCCCGAGCTATGCAGGACTCTGTGATGCCATGAAGCCAAATTGATGGAG 1673
304 tLysValAspIleAsnLeuGlnLysValAspIleAspGlnCysSerSera 321
1674 GAACTTTTGGAGTCCCTGATGAAACCAATTTTACTGGGGTTTCTGGAG 1723
321 spGlyTrpPheSerGlyThrHisLysCysHisLeuAsnAsnSerGluCys 337
1724 ATACGATCTTATTCGATGAGATGGAGACTCTCCAGGAAGGATGAATA 1773

```


[illegible][illegible]

```
3349 ACCACCGCGCGGC.....TCGGCCAGCGGACGAGCAGA 3386
866 uAlaGluSerThrGluSerValProLeuValCysLysSerAlaSerAlaH 883
3387 TGTGCGTGCCTGCACCTCGGACCGTGTGGCGGACGAGCTCTCTG.... 3432
883 isAsnLeuSerSerGluLysLysThrGlyHisProArgThrSerMetLeu 899
3433 .....CAGGCTCCCTCATG 3447
900 GlnLysSerLeuSerValIleAlaSerAlaLysGluLysThrLeuGlyLe 916
3448 GAGCAGATCAGCAGTGTGTACCCGCTTC..... 3477
916 uAlaGlyLysThrGlnThrAlaGlyValGluGluArgThrLysSerGlnL 933
3478 .....ACGCCAACATCAGCGAG..... 3495
933 ysProLeuProLysAspLysGluThrAsnArgAsnHisSerAsnSerAsp 949
3496 .....CTCAACTCCCATG 3507
950 AsnThrGluThrLysAspProAlaProGlnAsnSerAsnProAlaGluG 966
3508 ATGCTGTCCACCGCGCGCCCGCCAGCGCGTGTGGCGCGCGCTGTCTC 3557
966 luProArg...LysProGlnLysSerGly..... 974
3558 GTCCCTACCTGATCCCAAGAGATCCAGTGTCCCAAGACCATGACGACCT 3607
975 IleMetLysGlnGlnArgValAsnProThrThrAlaAsnSerAspLeuAs 991
3608 TTGCGGAATCAGCCTCTGCGCGGCTCGAAGTCAAGTCAAGCGGCGCTCAG 3657
991 nPro.....GlyThrThrClnMetLysAspAsnP 1001
3658 CCGCGCGGAGCGGCGGCGGCTGGGACGCGCGCGGAGAGCCCGC 3707
1001 heAspIleGlyGluValCysPro.....TrpGluValTyrAsp 1013
3708 GCGCGGTCCGAGGCTGCGCGCGCCAGCCAGACCTGGAGAGCTGTGTGG 3757
1014 LeuThrProGlyProValProSerGluSerLysValGlnLysHisValSe 1030
3758 CTCTACCGCGCGCTCCCTTCAGAGACTCGGTGGACTCGGGGAGCACA 3807
1030 rIleValAlaSerGluMetGluLysAsnProThrPheSerLeuLys.... 1045
3808 ACCCCCAACTCGCCAGTGTCCGAGTGGCCCTCTGTATCCCGTGTCTCC 3857
1046 .....GluLysSerHisHisLysProLysAla 1054
3858 CAATATCACACTCTTATCATAGAGATTACACTCAGAGCTCTCTCGTGT 3907
1055 AlaGluValCysGlnGlnSerAsnGlnLysArgIleAspLysAlaGluVa 1071
3908 TGTGAA..... 3913
1071 lCysLeuTrpGluSerGlnGlyGlnSerIleLeuGluAspGluLysLeuL 1088
3914 .TGTCCCTGGAAGACACGC..... 3931
1088 euIleSerLysThrProValLeuProGluArgAlaLysGluLysGly 1104
3932 .....CGGCTGGCGCTGCG..... 3946
1105 GlyGlnProArgAlaAlaAsnValCysAlaGlyGlnSerGluGluLeuPr 1121
3947 .....GAGCGGAGCC 3957
1121 oProLysAlaValAlaSerLysThrGluAsnGluAsnLeuAsnGlnIleG 1138
```

```
3958 CCCG.....TGTTCCACACACACAAT.....G 3980
1138 lYHisGlnGluLysLysThrSerSerGluGluAsnValArgGlySer 1154
3981 GCACGATAGTCGCTGTACGGCCAGGGGAGAGATGCCAAGGGCACC 4030
seq_name: /cgn2_5/ptodata/2/ina/5A_COMB.seq:US-08-486-270-9
seq_documentation_block:
; Sequence 9, Application US/08486270
; Patent No. 5807689
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; APPLICANT: Johnson, Edwin C.
; APPLICANT: Hess, Stephen D.
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,270
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/072,574
; FILING DATE: 04-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: FP41 9772
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9192
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4181 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 370..4008
; OTHER INFORMATION: /product= "HUMAN MGLUR5B"
; OTHER INFORMATION: /note= "Variant of MGLUR5A with 96 base pair
; insertion between nucleotides 2998 and 2999."
US-08-486-270-9
alignment_scores:
Quality: 208.00 Length: 1102
Ratio: 0.443 Caps: 43
Percent Similarity: 42.650 Percent Identity: 18.240
alignment_block:
US-09-775-181-2 x US-08-486-270-9 ..
Align seg 1/1 to: US-08-486-270-9 from: 1 to: 4181
```

219 GluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHi 235
 1366 GATGTCAAGTGGTTTGATGATTATCTGAAGCTCCGGCCAGAAACAA 1415
 235 sArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrpArgA 252
 1416 CCACCGAAACCTT.....TGGTTTCAAGAATTTTGGCAGC 1450
 252 rgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSerPro 268
 1451 ATCGT.....TTTCAGTGGCCAGCTGGAA 1473
 269 TyrLeuGluCysGluAsnGlySerTyr..... 277
 1474 GCGTTTCCACAGGAGAACAGCAATACAAACAGAGCTTGCAATAGTCTCT 1523
 278LysProGlyTrpLeuVal. 283
 1524 GACTCTGAACACACATCATGTTTCAGGATTCCAAAATGGGATTTGTGATCA 1573
 284 ..ThrLeuSerSerAlaIleTyrGlyLeu.....GlnProAsnLeu 296
 1574 ACGCATCTATTTCGATGGCTATGGCTCCACACATGCAGATGTCCCTC 1623
 297 ValProGluPheArgGlyVal.....Me 304
 1624 TGCCAGGCTATGCAGGACTCTGTGATGCCATGAAGCCAATTTGATGGACG 1673
 304 tLysValAspIleAsnLeuGlnLysValAspIleAspGlnCysSerSerA 321
 1674 GAACTTTTGGAGTCCCTGATGAAGAACCAATTTTACHGGGTTTCTGGAG 1723
 321 spGlyTrpPheSerGlyThrHisLysCysHisLeuAsnAsnSerGluCys 337
 1724 ATACGATCCTATTTCATGAGAAATGGAGACTCTCCAGGAAAGTATGAATA 1773
 338 MetProIleLysGlyLeuGlyPheValLeuGlyAlaTyrGluCysIleCy 354
 1774 ATGAATTCAAGGAATGGGA..... 1794
 354 sLysAlaGlyPheTyrHisProGlyValLeuProValAsnAsnPheArgA 371
 1795 .AAGATTACTTTGATTATATCAAGTTTGAAGTTGGACAATGGGAAT 1843
 371 rgArgGlyProAspGlnHisIle.....SerGlySerThrLys 383
 1844 TAAAAATGGATGATGATGAAGTATGTCTCAAGAAAGCAACATCATCAGA 1893
 384 AspValSerGluGlu..... 388
 1894 TCTGTGTGCAGTGAACCATGTGAGAAAGGCCAGATCAAGTGTGCGCAA 1943
 389AlaTyrValCysLeuProCysArgGluGly... 398
 1944 GGGAGAAGTCACTGTTGTTGGACCTGTACACCTTTGTAAGGAGAAATCAGT 1993
 399CysProPheCys..... 402
 1994 ATGTCTTTTGATGATACATCATCAAGGCTGCCAACTGGGTCTTGGCCC 2043
 403 AlaAspasp...SerProCysPheValGlnGluAspLysTyrLeuArg.. 417
 2044 ACTGATATCTACAGGTTGTGATTCATCCCAAGTACAGTATCTTCGATG 2093
 418LeuAlaIleLeuSerPheGlnGlyLeuCysMetL 429
 2094 GGGTGACCTTGAACCAATTCAGCTGTGGTGTTCCTGCTTGGCCCTCC 2143
 429 euLeuAspPheValSerMetLeuValTyrHisPheArgLysAlaLys 445
 2144 TGGCCACCTGTTTGTACTGTAGTCTTCATCATTTTACCGGTGATACACCA 2193
 446 SerIleArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySe 462

2194 GTAGTCAAGTCTCAACAGCAAGAACTCTGTACATATTATCTCTGTCGCAT 2243
 462 rLeuLeuLeuTyrPheProValIleLeuTyrPheGluProSerThrP 479
 2244 CTGCTGGGCTACTTATGTACCTTCTGCCCTCATTTGCCAAGCCCAACAGA 2293
 479 heArgCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrVal 495
 2294 TTTACTGTACTTACAGAGAATTGGCATTGTCTCTCTCCCGCAGCCATGAGC 2343
 496 TyrGlyThrValThrLeuLysLysHisArgValLeuLysVal..... 509
 2344 TACTCAGCCCTTGTACAAAGACCAACCGTATTGCCAAGGATCTGGCTGG 2393
 510PheLeuSerArgThrA 515
 2394 CAGCAAGAAGATCTGTACCCCAAGCCAGATTCATGATGAGCCCTGTG 2443
 515 laGlnArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAla 531
 2444 CCCAG..... 2448
 532 ValIleLeuValValPheTrpPheLeuIleGlyTrpThrSerSerVa 548
 2449CTAGTGATTGCTTTCATTCTCATTA..... 2472
 548 lCysGlnAsnLeuGluLysGlnIleSerLeu..... 558
 2473 TGCATCCAGTTGGCATCATCGTTGGCCCTTTTATTAATGAGGCTCCTG 2521
 559 ..IleGlyGlnGlyLysThrSerAspHisLeuIlePheAsnMetCysLeu 574
 2522 ACATAATGCATGACTACCAAGCATTCGAGAAGTCTACCTGATCTGTAAC 2571
 575 IleAspArgTrpAspTyrMetThrAlaValAla...GluPheLeuPheLe 590
 2572 ACCACCAACCTAGGAGTTGTCACTCCACTTGGAAACAATGGATTGTGAT 2621
 590 uLeuTrpGlyValTyrLeuCysTyrAlaValArgThrValProSerAlap 607
 2622 TTTGAGCTGCACCTTCTATCGGTTCAAGCACCAAGAAATGTTCCAGCTAACT 2671
 607 heHisGluProArgTyrMetAlaValAlaValHisAsnGluLeuIleIle 623
 2672 TCCCGAGGCCAAGTATATCGCTTCAATGTACACGACCTGCAATTATA 2721
 624 SerAlaIlePheHisThrIleArgPheValLeuAlaSerArgLeuGlnSe 640
 2722 TGCGTAGCTTTTGTTCCAATCTACTTT.....GGCAG 2753
 640 rAspTrpMetLeuMetLeuTyrPheAlaHisThrHisLeuThrValThrV 657
 2754 CAACTCAAAATCATCACCATTGTTTCTCGGTACAGCCATCATCTGCGCCACAG 2803
 657 alThrIleGlyLeuLeuLeuProLys.....PheSerHis 669
 2804 TGGCCCTAGGCTGCATGTTTGTGCGAAGGTGTACATCATCTACCGTGTGCGCAT 2853
 670 SerSerAsnAsnProArgAspAspIleAlaThrGlu...AlaTyrGluAs 685
 2854 CCAGAGAGAACCTGGCGCAGCGCTTCACCACATCTACCGTGTGCGCAT 2903
 685 pGluLeuAspMetGlyArgSerGlySerTyrLeu.....AsnSerSerI 700
 2904 GCATGTAGGGATGGCAAGTCAATCTCCGAGCCAGCAGATCCAGCAGCC 2953
 700 leAsnSerAlaTrpSerGluHisSerLeuAspProGluAspIleArgAsp 716
 2954 TAGTCAACCTGTGGAAGAGAAGGGGCTCTCTGGGAAACCTTTAAGGTAC 3003
 717 GluLeuLysLysLeu.....TyrAlaGlnLeuGluIleTyrLysAr 730

```

3004 AAGACAGGAGACTGCCCCAGCACAGTCCGGAATAGAGTGTTCACCCC 3053
730 gylslyMetile.....ThrAsnA 737
    |||
3054 CAAAGGAGTATGGGAATGGTGGGAGAGCAACAATGACAGTTCCTCAATG 3103
737 snProHisLeuGlnLysLysArgCysSerLysLysGlyLeuGlyArgSer 753
    |||
3104 GAAATCCGTACGTGGGCCAGAAATGAGAGAGCAGCGGGGAGCAC 3153
754 IleMetArgArgIleThr.....GluilePro..... 762
    |||
3154 CTGTGGCAGCGCTGTCCATCCATCCATCACACAGAAAGAACCCCAACCA 3203
763 .....GluThrValSerArgGlnCys 770
    |||
3204 AACGGCGTCATCAAGCCCTTCCCAAGAGCAGCGAGACCGTGGCGTGG 3253
770 erLysGluAspLysGluGlyAlaAspHisGly.....Thr 781
    |||
3254 GCGCTGGCGCTGGCGAGCGCGGAGCGCTGGGGCGTGGGGGCCACGGGC 3303
782 AlaLysGlyThrAlaLeuIleArgLysAsnProProGluSerSerGlyAs 798
    |||
3304 GTGCGGGCTGGCAGCGCCCGCCGCGCGCGCGCGAGTCCCCAGACGC 3353
798 nThrGlyLysSerLysGluGluThrLeuLysAsnArgValPheSerLeuL 815
    |||
3354 CGGCCCAAGGCG..... 3366
815 yLysSerHisSerThrTyrAspHisValArgAspGlnThrGluLuser 831
    |||
3367 .....CTGTATCATGTGGCGGAGCGCTGAG 3390
832 SerSerLeuProThrGluSerGlnGluGluThrThrGluAsnSerTh 848
    |||
3391 GAGCACTTCCCGCGCGCGCGCGCGCGCTCACCCTGCCCATCAGCAC 3440
848 rLeuGluSerLeuSerLysLysLysLeuThrGlnLysLeuLysGluAspS 865
    |||
3441 GCTGAGCCACCGCGCGGC.....TCGGCCAGCCGCGCAGGAG 3478
865 erGluAlaGluSerThrGluSerValProLeuValCysLysSerAlaSer 881
    |||
3479 ACGATGTGCGCTGCTGCTGCTGCGAGCGCTGCGCGCAGCAGCTCTCG 3528
882 AlaHisAsnLeuSerSerGluLysLysThrGlyHisProArgThrSerMe 898
    |||
3529 .....CAGGGCTCCCT 3539
898 tLeuGlnLysSerLeuSerValIleAlaSerAlaLysGluLysThrLeuG 915
    |||
3540 CATGGAGCAGATCAGCAGTGTGGTCCACCGCTTC..... 3573
915 lLeuAlaGlyLysThrGlnThrAlaGlyValGluGluArgThrLysSer 931
    |||
3574 .....ACGGCCAACATCAGGAG..... 3591
932 GlnLysProLeuProLysAspLysGluThrAsnArgAsnHisSerAsnSe 948
    |||
3592 .....CTCAACTC 3599
948 tAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsnProAla 964
    |||
3600 CATGATGCTGTCCACCGCGCGCGCGCGCGCGCTGCGCGCGCGCTCT 3649
965 GluGluProArg....LysProGlnLysSerGly..... 974
    |||
3650 GCTCTGCTTACCTGATCCCAAGAGATCCAGTTGCCACGACCATGACG 3699
975 ....IleMetLysGlnGlnArgValAsnProThrThrAlaAsnSerAspL 990
    |||
3700 ACCTTTGCCGAATCCAGCTCTGCCGCCATCCGAAGTCCAGGGCGGCGC 3749

```

```

990 euAsnPro.....GlyThrThrGlnMetLysAsp 999
    |||
3750 TCAGCCCGCGGAGGCGGCGAGCGGTGGGACGCGCGCGGAGAGCC 3799
1000 AsnPheAspIleGlyGluValCysPro.....TrpGluValTy 1012
    |||
3800 CGCGCGCGGTCCCGAGGCTGGCGCCCAAGCAGACCTGGAGGAGCTG 3849
1012 rAspLeuThrProGlyProValProSerGluSerLysValGlnLysHisV 1029
    |||
3850 GTGGGTCTCACCCCGCGCTCCCTTCAGAGACTCGGTGGACTCGGGAG 3899
1029 alSerIleValAlaSerGluMetGluLysAsnProThrPheSerLeuLys 1045
    |||
3900 CACAACCCCAACTCGCCAGTGTCCGAGTCCGCCCTCTGTATCCCGTCT 3949
1046 .....GluLysSerHisHisLysProLy 1053
    |||
3950 CTCCAAATATGACACTCTTATCATTAAGAGATTACACTCAGAGCTCCTCG 3999
1053 sAlaAlaGluValCysGlnSerAsnGlnLysArgIleAspLysAlaG 1070
    |||
4000 TCGTTGTGAA..... 4009
1070 luValCysLeuTrpGluSerGlnGlyGlnSerIleLeuGluAspGluLys 1086
    |||
4010 .....TGTCCTTGAAAGCACGC..... 4027
1087 LeuLeuIleSerLysThrProValLeuProGluArgAlaLysGluLys 1103
    |||
4028 .....CGGCTGCGCGTGGC..... 4042
1103 nGlyGlyGlnProArgAlaAlaAsnValCysAlaGlyGlnSerGluGluL 1120
    |||
4043 .....GAGCCGA 4049
1120 euProProLysAlaValAlaSerLysThrGluAsnGluAsnLeuAsnGln 1136
    |||
4050 GCCCCCG.....TGTTCCACACACACACAT..... 4075
1137 IleGlyHisGlnGluLysLysThrSerSerSerGluGluAsnValArgL 1153
    |||
4076 ...GCAAGCATAGTCGCTGCTTACGCCCGGAGGAGATGCCAAGGG 4122
1153 ySer 1154
    |||
4123 CACC 4126

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: November 1, 2001, 10:01:42 ; Search time 40.53 Seconds
(without alignments)
2283.545 Million cell updates/sec
Title: US-09-775-181-2
Perfect score: 6382
Sequence: 1 MGAWAYPELLICLLLAQLGLG.....LSANKIAPKKEIWDSEFKV 1215
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	279.5	4.4	433	2	T21989	hypothetical prote
2	225	3.5	5327	2	T13564	microtubule-associ
3	220.5	3.5	1199	2	A41939	G protein-coupled
4	210	3.3	1171	2	A42916	metabotropic gluta
5	210	3.3	1180	2	JC2132	metabotropic gluta
6	204.5	3.2	6642	2	T29757	protein UNC-89 - C
7	201.5	3.2	1200	2	A46194	neurofilament prot
8	199.5	3.1	1212	2	JC2131	metabotropic gluta
9	194.5	3.0	3488	2	T34418	hypothetical prote
10	190.5	3.0	734	2	B42680	nucleolus-cytoplas
11	190	3.0	1390	2	T14004	trfa protein - sli
12	188	2.9	1072	2	A86827	hypothetical prote
13	187	2.9	1046	2	T29776	hypothetical prote
14	186	2.9	522	2	C96608	hypothetical prote
15	183	2.9	852	2	T06310	hypothetical prote
16	183	2.9	1560	2	T42727	proliferation pote
17	181	2.8	1358	2	A29360	SIR4 protein - yea
18	179.5	2.8	1230	2	T22458	hypothetical prote
19	178.5	2.8	1020	2	T29108	hypothetical prote
20	177.5	2.8	729	2	S68191	triadin - human
21	177.5	2.8	3924	2	S37431	ankyrin 2, neuroa
22	177	2.8	532	2	T06029	hypothetical prote
23	177	2.8	667	2	A40713	cylicin I - bovine
24	177	2.8	1165	2	T16420	hypothetical prote
25	176	2.8	2094	2	S33124	tpr protein - huma
26	175.5	2.7	1837	2	T41023	probable nuclear p
27	174.5	2.7	1131	2	A49393	activator 1 large
28	174	2.7	1020	1	OFRUH	neurofilament trip
29	174	2.7	1558	2	T29253	hypothetical prote

30 173 2.7 919 2 F81998 ribonuclease E (EC
31 173 2.7 990 2 I51618 nucleolar phosphop
32 172.5 2.7 706 2 A45990 junctional sarcopl
33 172.5 2.7 1079 2 I59362 calcium/polyvalent
34 172.5 2.7 1218 2 S71376 glutamate receptor
35 172 2.7 1310 2 T40135 probable involveme
36 171.5 2.7 971 2 T19431 hypothetcal prote
37 171 2.7 1359 2 T34036 hypothetcal prote
38 170.5 2.7 879 2 JC7160 metabotropic gluta
39 170 2.7 904 2 T22457 hypothetcal prote
40 170 2.7 2218 2 B84683 hypothetcal prote
41 169.5 2.7 598 2 B40713 cylicin I - human
42 169.5 2.7 916 2 A27864 neurofilament trip
43 169.5 2.7 919 2 F81225 ribonuclease E NM8
44 169.5 2.7 2722 2 T20532 hypothetcal prote
45 169.5 2.7 5138 2 B96695 hypothetcal prote

ALIGNMENTS

RESULT 1
T21989
hypothetical protein F39B2.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T21989
R:Dobson, R.
submitted to the EMBL Data Library, March 1997
A:Reference number: Z19498
A:Accession: T21989
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-433 <WIL>
A:Cross-references: EMBL:Z92834; PIDN:CAB07388.1; GSPDB:GN00019; CESP:F39B2.8
A:Experimental source: clone F39B2
C:Genetics:
A:Gene: CESP:F39B2.8
A:Map position: 1
A:Introns: 137/2; 249/2; 287/3; 338/3
C:Superfamily: Caenorhabditis elegans hypothetical protein F39B2.8

Query Match 4.4%; Score 279.5; DB 2; Length 433;
Best Local Similarity 28.7%; Pred. No. 2.7e-08;
Matches 102; Conservative 57; Mismatches 129; Indels 67; Gaps 13;
QY 456 ETILFGSLLLYFPVILY-----PEP-----STFRCL---LRWRLGFRATVYG 497
DB 70 EHFSLTAVALKFPLQLAHFFANHSKDFDQLLDKSSQQIPDFDVSEARRTHFYGLQL 129
QY 498 TVTLKLRVILKVFSLTAQRIPYMTGGRVYMRMLAVILLVVFVFLIGTWSSVCQNLKQIS 557
DB 130 NCTGNHNRNLQYRVKKAQVHS-VREQDMUKYLAAMLATITGLMAWT----- 176
QY 558 LIQGGKTSOHLIFN-----MCLIDRWDMYMTAVAEFLFMGVLYCYAVRTVPSPAFHPRYM 613
DB 177 ---VGSWGDALTWRTPQCLMGQWHVWGHGYELLFLLYAVRLCYKARN--SDWLERWQF 231
QY 614 AVAVHNELIISAIFHTIRFVLASRLQSDWMLMYFAHTHTLVTTVITGLLIPKFSHSSNN 673
DB 232 TVAVCLEAVITLMANILIRYSIRNSGRADTLFIVSFVHLQLTVSVINIVVAPKFIYLSNGE 291
QY 674 PRDDIAEAYEDELDMGRSG-----SYLNSSINSANWSE-HSLDPEDIRDELKKL 721
DB 292 PSRRSMTLG-----GHSGRAHPSLAKURLNLTGTIDFAEVPIDMNPEDIRAEUKRV 344
QY 722 YAOLETYKRKKMTNNPHLQKRCSSKGLGRS---IMRRITEIPETVSRQCSKEDK 774
DB 345 YTOLRWYKLNLYQDNPHLSKREGKKWSKDKNTKATRIS-IP-----SCSPQTK 393

RESULT 2

Db 1022 -----PQPPPPQPPQPK--SLMDLOQVTVNFGSGIPDPHVLAVAGTGP- 1065

QY 997 MKDNFDIGVCPWEVYDLTPGPVPE--SKVQKHVSIVASEMEKNPTFSLKEKSHHKPKA 1054

Db 1066 -----GN-----SLRSLYPPPPPPHQLMLPLHLSTFQEEISIPGCEIDDDDS----- 1108

QY 1055 AEVCOOSNOKRIDAEVCLWESOG---QSTLEDEKLLISKTVPVLPRAKEENGCGQPRAN 1111

Db 1109 -----ERFKLOEFVYEREGNTEDELEEDLPTASKLTDPEDSPALTPPSPFRDS 1159

QY 1112 VCAGQSEELPP--KAVASCTENENLNOIGHOEKKTSSS 1147

Db 1160 VASGSSVPSSPVSESLCPTPPNTYASVILRDYKQSSS 1197

RESULT 4

A42916

Species: Rattus norvegicus (Norway rat)

C: Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999

R: Abe, T.; Sugihara, H.; Nawa, H.; Shigemoto, R.; Mizuno, N.; Nakanishi, S.

J. Biol. Chem. 267, 13361-13368, 1992

A: Title: Molecular characterization of a novel metabotropic glutamate receptor mGluR5

A: Reference number: A42916; MUID: 92317054

A: Accession: A42916

A: Status: preliminary

A: Molecule type: mRNA

A: Residues: 1-1171 <ABE>

A: Cross-references: GB:D10891; NID:q220813; PIDN:BAA01711.1; PID:dl002186; PID:q220814

A: Experimental source: brain

A: Note: sequence extracted from NCBI backbone (NCBIN:107749, NCBI:107750)

C: Keywords: G protein-coupled receptor; transmembrane protein

Query Match 3.3%; Score 210; DB 2; Length 1171;

Best Local Similarity 19.1%; Pred. No. 0.00075;

Matches 179; Conservative 117; Mismatches 397; Indels 244; Gaps 35;

QY 219 ETEWFHGLRRKWRPHLHRRGNQGRGLGHSWR-----RKDGLGGDKSHFKWSPPYLECE 273

Db 332 DVKWFDDYLLKRLPEINLRNP-----WFOEFWQHRFOCRLEGFAQENSKYNTK-----CN 381

QY 274 NG-----SVKPGMLV--TLSSATYGL---OPNLVPEPRGV-----MKVDINLQ 311

382 SSUTLTHRHVQDSKMGFVINAISYMAVGLHNMQMSLCPGVAGLCDAMKPIDGRKLLDSLM 441

312 KVIDQCSSDQWFGSGTHKCHLNSECMPIKGLGVGLGAYECICKAGFYHPGVLPV----- 366

442 KTNFTGVSGDMILFDENGSDSPGRYEIMNFKEMG---KDYPDYINVGSWDNGELKMDDEV 498

QY 367 ----NNFRRRGPDQHI--SGSTKDVSEE-----AYVCLPCREG-----CPFC-----A 403

499 WSKKNILIRSCPECKEQIKVIRKGEVSCCWTCTPCRENEYVDEYTCACQLGSWPT 558

QY 404 DD--SPCFVQEDKYLR-----LAIISFGQCLMLDLFVSMVLVYVHFRAKASIRASGILLLE 456

559 DDLTGCDLIPVQVLRMGDPPEIAAVVAFACGLLATLVIFVIIFIVITPTPVKSSRELCE 618

QY 457 TILFGLSLLFPVIVLYFESTPRCILLRWARLLGFATVGTVTLKLRHLKV----- 509

619 IILAGICGLVCTFCLIAKPKQICYLQRIIGLISFAMSYSALVTKTNRIARLAGSKK 678

QY 510 -----FLSRTAQRIPTYMTGGRVMRMLAVILLVWFELWFLGWTSSVCQNLKEOISL----- 558

679 ICTKKPRFMSACAQ-----LVIAFILI-----CIQLGIIVAFIME 714

QY 559 ----IGOKTSDHLLFNKCLDRWDYMTAVA--EFLFLMWVYLCYAVRTVPSAFHEPRYMA 614

715 PPDIMHDYPSIREVYLCNTNLGVVTPGLYNGLLILSCFTYAFKTRNVPANENAKYIA 774

QY 615 VAVHNELIISAIFHTFRFVLASRLQSDWMLMLYFAHTLTVTIVTIGLLIIPK-----FSSHS 670

Db 775 FTWYTTCTIWLAEVPIYF-----GSNYKIITMCFVSLSATVALGCMFVPKVIYIILAKP 828

QY 671 SNNPRDIIATE-AYEDELDMGRSGSYL--NSSINSASWSEHSLEDIPDIRDELKK----- 720

829 ERNVRSAFTTSTVVRMHVGDGKSSAARSSSSLVNLWKKRSGSGETLSSNGKSVTWAONE 888

QY 721 -----LYAQLBIYRKMKMITNNPHLOK---KRCSSKGLGLGRSIMRRITEIPETVSRQCS 770

889 KSTRGQHLWQRLSHVINKKENPNQTAIVKPPFKSTENRG-----PGAAAGGGS 936

QY 771 KEDKGAHGCTAGTALIRKNPPSSGNTGKSKEETLKNRVFSLKSKSHSYDHRVQDTEE 830

937 GPCVACA--GNACCTATGGPEPPDAGPKA-----LYDVAEA 970

QY 831 SSSLPTESQEEETSTLESLSGKKLTOKLKEDSEAESESVPLVCKSAHNLSEKK 890

971 EESFPAAARPRSPSPITSLHLAG-----SAGRTDDDDAPSLHSETAARSSSSQGS 1020

QY 891 TGHPRSTMLOKLSVITASAKETLGLAG-----KTOTAGVEER 928

1021 LMEQISVVTRFTANISELNSMMLSTAATPGPGTPTICSSYLLIPKEIQLPTMTTFAE-- 1078

QY 929 TKSQKPLPKDKETNRNHSNDTETKDPAPONSNAPEE-PRKPKQSGIMKQQRVNPTAN 987

1079 ---IQPLPAIEVT---GGAQGATGVSPAQETPTGAESAPGKPDLEELVALTPPSP----- 1127

QY 988 SDLNPQTQMKDNFDIGEVCPWEVYDLTPGPVPSSEK 1024

1128 -----FRDSVDSSGTTTPNSPVSESLCIPSSPK 1155

RESULT 5

JC2132

metabotropic glutamate receptor 5 A - human

C: Species: Homo sapiens (man)

C: Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 05-Jan-1996

R: Minakami, R.; Katsuki, F.; Yamamoto, T.; Nakamura, K.; Sugiyama, H.

Biochem. Biophys. Res. Commun. 199, 1136-1143, 1994

A: Title: Molecular cloning and the functional expression of two isoforms of human met

A: Reference number: JC2131; MUID: 94197696

A: Accession: JC2132

A: Molecule type: mRNA

A: Residues: 1-1180 <MIN>

C: Comment: This protein is coupled to guanine nucleotide binding proteins.

C: Keywords: glycoprotein; neurotransmitter; receptor; transmembrane protein

F: 580-604/Domain: transmembrane #status predicted <TM1>

F: 617-637/Domain: transmembrane #status predicted <TM2>

F: 644-664/Domain: transmembrane #status predicted <TM3>

F: 694-714/Domain: transmembrane #status predicted <TM4>

F: 738-759/Domain: transmembrane #status predicted <TM5>

F: 773-794/Domain: transmembrane #status predicted <TM6>

F: 803-827/Domain: transmembrane #status predicted <TM7>

Query Match 3.3%; Score 210; DB 2; Length 1180;

Best Local Similarity 19.1%; Pred. No. 0.00076;

Matches 179; Conservative 120; Mismatches 400; Indels 240; Gaps 34;

QY 219 ETEWFHGLRRKWRPHLHRRGNQGRGLGHSWR-----RKDGLGGDKSHFKWSPPYLECE 273

Db 333 DVKWFDDYLLKRLPEINLRNP-----WFOEFWQHRFOCRLEGFPQENSKYNTK-----CN 382

QY 274 NG-----SVKPGMLV--TLSSATYGL---OPNLVPEPRGV-----MKVDINLQ 311

383 SSUTLTHRHVQDSKMGFVINAISYMAVGLHNMQMSLCPGVAGLCDAMKPIDGRKLLDSLM 442

QY 312 KVIDQCSSDQWFGSGTHKCHLNSECMPIKGLGVGLGAYECICKAGFYHPGVLPVNNFR 371

443 KTNFTGVSGDTILFDENGSDSPGRYEIMNFKEMG-----KDYPDYINVGSWDNGEL 492

QY 372 RGPDQHI-----SGSTKDVSEE-----AYVCLPCREG-----CPFC 402

Query Match 3.2%; Score 201.5; DB 2; Length 1200;
Best Local Similarity 23.4%; Pred. No. 0.0023;
Matches 131; Conservative 89; Mismatches 248; Indels 93; Gaps 25;

709 LQPEDIRDELKLLIYAOL--EYVKRKKMTNNPHLQKRCSKGLGRSTIMRRITTEIP-- 763
::
Db 442 LEGESRIELVHPWTIGTREAAYRPDLITNGKSASDDSSKDGTVRAKSVSPDVVAETK 501
764 --TVSRQC--SKREDKEGADHGTAAGT-ALIRKNPPSSGNTGKSEETIKNRYFSLKSH 818
502 LTTSTSYCGDEADDGCKSDSDTHTEAEEHTRADSDADTGTLGVLDVEESVLKSEED 561
819 SYIDHRVDQTQESSLIPTSQEBEETTSTLSLKKLTQKLKED-SRAESTESVPLVC 877
562 KSVKRDDEDEEENDQTESVDE--DEADIEK-SOKSESDDTKESKAESRAESVKAAT 618
878 KSAAS--AHNLSSPKKHCHPRTSMLOKSLVSIAAKEKTGLACKTOTAGVERTKSOKPL 935
619 PSEDKAESVKSERSSITSETAFIDKSPKSPGS-KRDTLFQSDKVTPSVLLERAMSLQA 677
936 PKDKETRNIHSNDNET---KDPAA-PONS-----NPAEPKPQSGIMKQORVPN 983
678 PTEKPL-----SDTIKSPVWSEPALSIVSSVRSGAETSFAERTSPIGSGSEKSAK-SP 730
984 TTANSLNPGTQMKNDFDIGEVCPMEVYDLTPGPVPE-----SKVKHVSIVASEM 1036
731 VRSEATKSP-----VSEKSAGSKSPVPSEALSPVLSEMSVH-STAMSHT 776
1037 EKNPTFSLEKESHHKPKAE-----VCOOSNQKRIDKAECVCLNESQGOSILEDK 1086
777 SRSPITAS-EKSVKSPTHSERTASPTAKSPIIMEPAKSPKDESEKELSPERSSEVMGKS 834
1087 LLIS----KTPVLPILERA-KSENGQPRAANVCAGOSELPP-----KAVASKTENEN 1133
835 QITSSAKSPVSEKADSEKATSPTPEKVVDESARSPTHSDRSSEKSAKSPMTSDH 894
1134 LNOIGHOEKHTTSSEENVRGSYNSNNFOOPLTS-----RAEVCWPWFETTPAOPNAGRS 1187
895 IKSPIDFEKAES---EKSLAKYSKSDHDKSPVPSEKAESKARSVPSEKAESEKSARS 951
1188 --VALPASSALSANKTAGPRK 1206
952 PVTSEPAKSPVPSEKAESAEEK 972

RESULT 8
JC2131
metabotropic glutamate receptor 5 B - human
C:Species: Homo sapiens (man)
C:Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 05-Jan-1996
C:Accession: JC2131
R:Minakami, R.; Katsuki, F.; Yamamoto, T.; Nakamura, K.; Sugiyama, H.
Biochem. Biophys. Res. Commun. 199, 1136-1143, 1994
A:Title: Molecular cloning and the functional expression of two isoforms of human metabo-
A:Reference number: JC2131; MUID:94197696
A:Accession: JC2131
A:Molecule type: mRNA
A:Residues: 1-1212 <MIN>
C:Comment: This protein is coupled to guanine nucleotide binding proteins.
C:Keywords: glycoprotein; neurotransmitter; receptor; transmembrane protein
F:580-604/Domain: transmembrane #status predicted <TM1>
F:617-637/Domain: transmembrane #status predicted <TM2>
F:644-664/Domain: transmembrane #status predicted <TM3>
F:694-714/Domain: transmembrane #status predicted <TM4>
F:738-759/Domain: transmembrane #status predicted <TM5>
F:773-794/Domain: transmembrane #status predicted <TM6>
F:803-827/Domain: transmembrane #status predicted <TM7>

Query Match 3.1%; Score 199.5; DB 2; Length 1212;
Best Local Similarity 18.2%; Pred. No. 0.003;

Query Match	3.1%	Score 199.5;	DB 2;	Length 1212;
Best Local Similarity	18.2%	Pred. No. 0.003;		

Matches	204;	Conservative	129;	Mismatches	381;	Indels	409;	Gaps	45
Qy	219	ETEFWHLRRKWRPHLHRRGPNQGRGLGHSWR	-----	RKDLGGDKSHFKWSPPYLECE	273				
Db	333	DKVWEDDYLLKLRPETNHRNP	-----	WFQEFWQHRFOCRLEGFPQENSKYNT	382				
Qy	274	NG-----SYKPGWLIV-TLSAIIYGL	-----	QNLNVPEFRGV	311				
Db	383	SSLTKTHHVQDSKMGFVINAIYSWAGLHNMQMSLCPYAGLGDAMKPIDGRKLLSML	442						
Qy	312	KVDIDQSSDGWFSGTHKCHLNNECEPIKGLGFVGLAYECICKAGVHPGCVLPVNNFR	371						
Db	443	KTNTGVSQDITLFDENGDSGRVEIWNFKDMG	-----	KDYFDYINVGSWMDGEL	492				
Qy	372	RGPDQHI-----SGSTKDVSEB	-----	AYVCLPCREG	402				
Db	493	KWDDDEVSKNSIIRSVCSPECEKGOIKVIRKEVSCCWTCTPCKENEYVDFEYTCCK	552						
Qy	403	-----ADD-SPCFVQEDKYLR	-----	LATISQGLCMLLDFVSMVLVVYHFRKAKSIRA	449				
Db	553	QLGSWPTDDLTCGDLIPVQYLRWGDPEPIAAVFAECLGALLATLFTVVFYIIRDTVPVKS	612						
Qy	450	SGLLLETLFGSLLLYFPVVLVFEESTPCILLRWALLGFATVGTVTLKLRHLVKV	509						
Db	613	SSRELXVILLAGICGLYCTCFCLIAKPKQIYCYLQRIIGLSPAMSYALSATKTNRIARI	672						
Qy	510	-----FLSRTAQIRPYMTGGVRMLAVILLVFWFLGWTSSVCONLEKQ	555						
Db	673	LAGEKKKICTKKPRFNACAO	-----	LVIARILI-----CIQLGII	708				
Qy	556	ISL-----IGOGKTSDHLFNMLIDRWDMYMTAVA	EFLLGLMWGLYCAVRTVPSAF	607					
Db	709	VALFIMEPPDITMHDYPSIREVYLICNTNLGVVPLGYVGLLILSCSTFYAFKTRNVPANF	768						
Qy	608	HEPYMAVAVINELIISAIFITRFLVASRLOSWMMLMYFAHHTLHVTVTIGLLLPK	666						
Db	769	NEAKYIAFTMTTTCIIWLAFPIYF-----GSNKIITMCFVSLSATVALGCMFVKV	822						
Qy	667	---FSSHNSNPDDIAYE-AYEDELDMGRSGSYL-	NSSTNSANSEHSLDPEDIRDELKK	720					
Db	823	YIILAKPERNVSAFTSTVVRMHVGDGKSSASRSSSLVNLKRGSSCETLR	877						
Qy	721	LYAOLEYIKRKMINTNPHLOKRCCKGLGRSITMRITEIPETVSRQCKEDKEGADHG	780						
Db	878	-----YKDRRLAQHKSEI	-----ECF-----	893					
Qy	781	TAKGTALIRKNPPSSNGTCKSKETLKNRVFSLLKSHSTYDHYDQDTESSSLPTESQE	840						
Db	894	TPKG-----SMGNGR-----	904						
Qy	841	EETTENSTLESGLKKTOKLKEDESEASTESVPLVCKSASAHNLSSEKKTGHGHPRTSMLQ	900						
Db	905	-----ATMSSSNKGSVTWAONEKSS	-----RGQHLW	930					
Qy	901	KSLSVIASAKETLGLAGTKTQTAGVEERTKSQKLPKDKET	-----NRNHSNS	948					
Db	931	QRLSIHINKKE-----NPNQTAVI-----	KFPKSTESRGLGPGRGAGSGAGVCAT	977					
Qy	949	-----DNTETKDPAPQ-----NSNPAEE	-----PRKPKSGIMKQORVNPPTTANS	988					
Db	978	GGAGCAGAGPGGPSPDAGPKALYDVAEEHPFAPAPRSPSPISLISHR	1028						
Qy	989	DLNPGTQMKDNFDIGEVCPHEVYDLPGPV	-----PSESKVQKHVSIVASEMEKN	1040					
Db	1029	---AGSASRTDD-----DVPSLHSEPVARSSSQGSLSMEQISSVVVTRFTANISELN	1076						
Qy	1041	TFSLKESKSHHKPKAAEVCQSQSNQKRIDKAEVCLWESQCQSTLEDEKLLISK	1095						
Db	1077	SMMLSTRAAPSPGVGAPLC	-----SSVLIPEKIQLPYMTTFAEIQ	1116					
Qy	1096	PERAKEENG-OPRAANVCAG-----QSEELDPKAVASKTENENLNQIHQEKTKTSSEEN	1150						
Db	1117	PLPAIEVTGGAPQAAGAAQAGDAARESPAAGPEAAAKPDLLEUVAL	1168						

QY 1151 VRGSSNNFQOPLTSRAEVCWPFEETPAQPNAGRSVALPAS 1193
| | | : : : : :
Db 1169 FRDSVDSGST-----TPNSPVSEALCIPSS 1194

RESULT 9
T34418
hypotheical protein F12F3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34418
R:Fulton, B.; Wohldmann, P.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid F12F3.
A:Reference number: Z21521
A:Accession: T34418
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3488 <FUL>
A:Cross-references: EMBL:U80022; PIDN:AAC25885.1; GSPDB:GN00023; CESP:F12F3.3
A:Experimental source: strain Bristol N2; clone F12F3
Genetics:
Gene: CESP-F12F3.3
Map position: 5
A:Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

Query Match 3.0%; Score 194.5; DB 2; Length 3488;
Best Local Similarity 20.9%; Pred. No. 0.024;
Matches 136; Conservative 81; Mismatches 228; Indels 205; Gaps 24;

QY 705 SEHSLDPEDI-----RDELKLYAQLEIYKRRKMTNNPHLQKRCCKGLGRSIMRRIT 759
| | | : : : : :
Db 933 SKKTTIDKVGATEPADETPK-----KKII-----KKTEKS---DSSISQSKA 973

QY 760 EIPETVSQCSKED-----KEGADHGTAAGTALIRKNPPSSGNTGSKKETL 807
| | | : : : : :
Db 974 TDEKVSQKQEQDEPTKPAVSETQMTVEADKSKQKQETDEKLKDAETAAKTQEADEKS 1033

QY 808 KNRVFSLKSHSTVDHVRDQTEESSLPTESSQ-----EEETTSTNLSLSG 854
| | | : : : : :
Db 1034 KLDQAQETKKVSDDAARKEKELNDKLLKLESEIATKASADKLLKEQQAQAKAAVEAA 1093

QY 855 KKLTKQ---LKEDSEAEATESVPLVCKSSAHLNLSSEKKTGHPTSMLOKSLVSIAAKE 911
| | | : : : : :
Db 1094 KKQEKDEQLKDTAAAS-----KKAEEKLELEKQ-----AQIKKAAGADAVKKQ 1139

QY 912 KTLG-----LAGKTQTAG-----VEERTK-----SOKPLPKDKE 940
| | | : : : : :
Db 1140 KELDEKNKLEANKKSAAGKLEIEESAASQTVVEEQAKLDAQTKAKTAETKOTLEKDEK 1199

QY 941 TNRNHSNDNPTETKDPAPONSNAEPRKPKOKSGIMQOORVNPPTTNSDLNPGTOMKDN 1000
| | | : : : : :
Db 1200 STK---ESESKEVDEKPKKKVLKKKTEKSDSS---ISQKSTSTVTVESAGPSESTQKV 1254

QY 1001 FDI-----GEVCPWEVYDLTPGPVPSKVKQKHVSIVASEMEKNPTFSLK- 1045
| | | : : : : :
Db 1255 ADAARKQKETDEKQKLEAEITAKKSAD-EKSKLEAESKLLKAAVEAAKQKQKDEQLK 1313

QY 1046 -----PKSHHKPKAAEVQCSQKQKRIDK----- 1068
| | | : : : : :
Db 1314 DTEAASKAAAEKLELEKQSHIKKAAEVDVAKKQKLEKEKQORLESEAAATKKADEKLEKLE 1373

QY 1069 -----AEVCLWE-----SQQSQSTLEDEKLLISKTPVLPERAKENGQOPRAANVCA 1114
| | | : : : : :
Db 1374 EOKKAAEIALIELQKQEKLAQOSRLEDE-----AKSAEKQKLESETKS 1420

QY 1115 GQSEELPPKAVASTENENLNQIGHQEKTSSEENVRGSSNNFQOPLTSRAEVCWP 1174
| | | : : : : :
Db 1421 KOTEEAPKESVDEKPKKKVL-----KKTEKSDSSISQKSKSAK-----TVDAAEETLS 1470

QY 1175 EF-----ETPAQPNAGRSVALPASSALSANKIA-----GPRKEBIWD 1211

Db 1471 DFNLVEKKTVQKVEQSPDESTSATIKRDPQAQKTEISKODDGEKKTTFD 1520
| | | : : : : :
RESULT 10
B42680
nucleolus-cytoplasm shuttle phosphoprotein - rat
N:Alternate names: Noppi40 protein B
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C:Accession: B42680; A42680; S27890; S27889; S30510; S30511
R:Meier, U.T.; Blobel, G.
Cell 70, 127-138, 1992
A:Title: Noppi40 shuttles on tracks between nucleolus and cytoplasm.
A:Reference number: A42680; MUID:92323542
A:Accession: B42680
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-734 <MEI>
A:Cross-references: GB:M94288
A:Experimental source: clone ptm6
A:Accession: A42680
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 32-180, 'Q', 181-734 <ME3>
A:Cross-references: GB:M94287; PIDN:AAA41718.1; PID:g205750
R:Meier, U.; Blobel, G.
submitted to the EMBL Data Library, May 1992
A:Description: Noppi40 shuttles on tracks between nucleolus and cytoplasm.
A:Reference number: S27889
A:Accession: S27890
A:Molecule type: mRNA
A:Residues: 32-734 <ME2>
A:Cross-references: EMBL:M94288; NID:g205751; PIDN:AAA41719.1; PID:g205752
A:Accession: S27889
A:Molecule type: mRNA
A:Residues: 32-180, 'Q', 181-734 <MEI2>
A:Cross-references: EMBL:M94287; NID:g205749; PIDN:AAA41718.1; PID:g205750
C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

Query Match 3.0%; Score 190.5; DB 2; Length 734;
Best Local Similarity 21.3%; Pred. No. 0.0048;
Matches 126; Conservative 92; Mismatches 239; Indels 135; Gaps 25;

QY 665 PKFSHSSNNP-----RDDIATEAYED---ELDMGR-----SGSYLN 697
| | | : : : : :
Db 95 PKVKLQNGPVAKKAKKETSSDSEDSSEEDKAAQVPTQKAAAPAKRASLPQHAGKAAA 154

QY 698 SSINSAWSEHSLDPEDIRDELKLYAQLEIYKRRKMTNNPHLQKRCCKGLGRSIMRR 757
| | | : : : : :
Db 155 KASSESSSESESESESEKDK-KKPVQKAVKQAKAVRPP---KKAESSESEDSSE- 209

QY 758-ITEIPETVSQCSKEDKEGADHGTAAGTALIRKNPPSSGNTGSKKETLKNRVFSLKS 817
| | | : : : : :
Db 210 -DEAPQT-----QKPKAAATAAKAPTRAQTKAPAKPGPPAKAQPKAANGKAGSSSS 260

QY 818 HSTVDHYRDQTEESSLPTESSLPTESSLPTESSLPTESSLPTESSLPTESSLP 876
| | | : : : : :
Db 261-----SSSSSDSDSESEKAAAPLKKTAQKQVAKAPVKVTAAPTQ----- 302

QY 877-CKSASAHNLSE-----KKTGHPTSMLOKSLVSIAAKEKTLGLAGKTQTAGVEE 927
| | | : : : : :
Db 303-KSSSDSDSDSESESESEKDKKPKKKAGPYSSVPPPSVSL-----SKKSVG----- 345

QY 928 RTYSQKPLPKDKDITNRNHSNDNPTETKDPAPONSNAEPRKPKQSGIMK-QQRVNPTTA 986
| | | : : : : :
Db 346---AQSPKAAAQOTQPADSSADSSSESD-----SSSEEEKTKPAKTVPVKTAPKAPVKK 397

QY 987 NSDLNPTCTQMKONFDIGEVCPWEVYDLTPGPVPSKVKQKHVSIVASEMEKNPTFSLKE 1046
| | | : : : : :
Db 398 KAE-----SSSDSDSDSDSEDEAPAKPVSAATKSPKPAVTPKPKPAKAVATPKQPA----- 449

```

Qy 1047 KSHHKPKAAEVCQSNQKRIDKAEVCLWESQGSQSTLEDEKLLISK-TPVLPERAKEENG 1105
      |||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
Db 450 GSGQKPOSRKADSSSEESSSE---BATKKSVTTPKARVTAKAAPSLPKAQAPRAG 506
      |||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
Qy 1106 QPRAANVCAGOSSE--LPKAVASK-----TENENLNOIGHOEKKTPTSSEENVRSYN 1156
      |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||:: ||
Db 507 DSSSDSESSSESEKTPPKPAKKKAAGAAYPKTPPKKAAASSSSSSSSE-----D 560
      |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||:: |||:: ||
Qy 1157 SSNNFOPLTSRAEVCWFETTPAQPNRGSVALPASSALSANKTAGRKEE 1208
      ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
Db 561 SSEEKKKPKSKA-----TP-KPQAGKANGVPASQ---NGKAGKESEE 599
      ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||

RESULT 11
T14004
trfA protein - slime mold (Dictyostelium discoideum)
Species: Dictyostelium discoideum
Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
Accession: T14004
R:Salto, J.; Adachi, H.; Sutch, K.
J. Biol. Chem. 273, 24654-24659, 1998
A:Title: Dictyostelium TRFA homologous to yeast Ssn6 is required for normal
A:Reference number: Z17852; MUID:98406112
A:Accession: T14004
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-11390 <SAT>
A:Cross-references: EMBL:AB009080; NID:d1228566; PID:d1034109; PIDN:BAA3314
C:Genetics:
A:Gene: trfA
A:Introns: 333/3; 364/3; 637/1

```

[illegible]

RESULT¹²
 A86827
 hypothetical protein yqfg [imported] - *Lactococcus lactis* subsp. *lactis* (strain IL140)
 C:Species: *Lactococcus lactis* subsp. *lactis*
 C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
 C:Accession: A86827
 R:Boilotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Eh
 Genome Res. In press, 2001
 A:Title: The complete genome sequence of the lactic acid bacterium.
 A:Reference number: A86625
 A:Accession: A86827
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1072 <STO>
 A:Cross-references: GB:AE005176; NID:g12724625; PIDN:AAK05715.1; GSPDB:GN00146
 A:Experimental source: strain IL1403
 C:Genetics:
 A:Gene: vafg

[illegible]

10

QY 1157 SSNNFOPLTTSRAEVCWPEFTTAPQNAKGRSVLPASSALS-ANKIAGRKEE 1208
Db 841 SSIOQAILES-----SKSSTNKRSSLSLIINSTSHQPNED 876

RESULT 13
T29776
hypothetical protein C50F2.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29776
R:Du, Z.; Le, T.T.
submitted to the EMBL Data Library, November 1996
A:Description: The sequence of C. elegans cosmid C50F2.
A:Reference number: 220684
A:Accession: T29776
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1046 <DUZ>
A:Cross-references: EMBL:U80445; PIDN:AAB37795.1; GSPDB:GN00019; CESP:C50F2.2
A:Experimental source: strain Bristol N2; clone C50F2
Genetics:
A:Gene: CESP:C50F2.2
A:Map position: 1
A:Introns: 72/2; 888/1; 940/1; 986/3

Query Match 2.9%; Score 187; DB 2; Length 1046;
Best Local Similarity 22.3%; Pred. No. 0.012;
Matches 122; Conservative 86; Mismatches 220; Indels 118; Gaps 26;

QY 676 DIATEAYEDELDMGRSG-----SYLNSSINSAMSEHSLSLDPEDI-----RDELKLYAQ 724
Db 138 EDEMLYLINQLETTNDGPTVYNNKATNCPEKSLKESGLLPPELVKKREKQEDQ 197

QY 725 -LEIYRKMMITNNPHLOKRCCKGLGHSIMRRITEIPETVSRQCSKEDKAGDHGTA 783
Db 198 KAVLEKRPVTKVQDHLKRSKSPIGRGQSDKDIISNKKRLNKREKEIDE----- 253

QY 784 GTALIRKNPSSNGTGSKEETLKNRV-FSLKKSHSTVDHVRDOTESSSLP-TESQEE 841
Db 254 -RSK TAKAP-----ISQVQFPMPAPQAPDAVASKVRQKQVDITPHPRP 299

QY 842 ETENSTLESGLKLTOKLKEDSEASTESVPLVCKSASAHNLSSEK-----TGHP--- 894
Db 300 SASRNSQKPVASKIQEK-KTEVRTKSSPTNPVINSQ-----PSQKKPADLTPHPKRP 353

QY 895 -----RTSMLOKSLSVIASAKEKTLGLAGKTOTAGVEERTK-SOKPLPKQKETNRN 944
Db 354 PITPETLPSRNSIQPAA--SKIQEKTEVIARSQKS--EASTKIAQKPSPTNPVNFK 409

QY 945 HNSDNT-----TKDAPQNSPAEEP-RKPOKSGIMKQORVNPPTANSDLNPGT 994
Db 410 QPSQPADLTPKPIPLASKSPITRSEVSTERTPKRAPIVMDDPKDSTA-KELPKEL 468

QY 995 TOMKONFDIGEYCPWEVYDLTPGPVPSKVKQHSIVASEMEKNPTFSLKESHKPKA 1054
Db 469 PRAKSHFEF-KVKPIEKDASPKTTSTS-----SASPPNPDPAMKQKVFVTS 517

QY 1055 ARVCOOSNQKRDKAECVLWESQGSILDEKLLISKTPVLPERAKENGGQOPRAANVCA 1114
Db 518 -----KKOK-----LTETKN-----KQKSTINESPV-----EQNSGQNEASNTK 554

QY 1115 GQSEEL-----PPKAVASKTENENLQIGHQEKTKSSSENVNRYSGYNSNNFQOPLTSRA 1169
Db 555 LKQELNKHVPKNPEAETPD-----RQESVOKSNEASATSKPSDPVETP--SD 605

QY 1170 EYCPWE 1175
Db 606 EYDPOE 611

RESULT 14
C96608
hypothetical protein F25P12.91 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C96608
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
Nelson, N.F.; Hughes, B.; Huiziar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C96608
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-522 <STO>
A:Cross-references: GB:AE005173; NID:g9954748; PIDN:ANG09099.1; GSPDB:GN00141
Genetics:
A:Gene: F25P12.91
A:Map position: 1

Query Match 2.9%; Score 186; DB 2; Length 522;
Best Local Similarity 19.7%; Pred. No. 0.0054;
Matches 96; Conservative 90; Mismatches 218; Indels 84; Gaps 15;

QY 709 LDP-----DIRDELKLYAQLEIYRKMMITNNPHLOKRCCKGLGHSIMRRITEIPET 764
Db 22 LDPKEGNEVMEVAKSIEKVRKAKDESSGSKKDKK--GKNVDEVKEDDD 79

QY 765 VSRQ-----CSKEDKGGADHG---TAKGTALIRKNPSSNGTGSKEETLKNRVFSLKKS 817
Db 80 DKKDKGKVMYKHEEG--HGDLEVKESDVKEHEKHKKGKKEHEELEEKEGKKKN 137

QY 818 HSTYDHRDOTEESLSLPTESQEEETTENTSTLESLSGKKLTOKLKEDSEASTESVPLVC 877
Db 138 KKEKDESGPEKNKADKKEKHEDVSQKEELEEDGKNNKKKKEDESGTEKKKKPKKE 197

QY 878 KSASAHNLSSEKKTGHPRTSMLOKSLSVIASAKEKTLGLAGKTOTAGVEERTKSQKPLPK 937
Db 198 KKQKEESKNEK-----KVKGKKEGKEDLEKEDEKK--KE 234

QY 938 DKETNRNHSNDNTETKDPAPQNSPAEPRKPKQSGIMKQORVNPPTANSDLNPGTQM 997
Db 235 HDETQEMKEKDKSKNNKKKDESCAEEKKKPKDEKDEKEDKELKKGKKGK 294

QY 998 KDNEDIGEYCPWEVYDLTPGPVPS-----KVKQHSIVASEMEKNPTFSL 1044
Db 295 PEKEDEGK--KTKHEDATEQEMDDEADHKEGKKKKKKKAKKKTVIDEVECK-----ET 348

QY 1045 KEKSHHKPAAEVCOOSNQKRDKAECVLWESQGSILDEKLLIS--KTPVL----- 1095
Db 349 KDKODDE--GETKQKNNKKEKSE-----KGEKVDKDKKENPLETEVMSRDIKLE 399

QY 1096 ---PPRAKEENGQOPRAANVAGQSEELPPKAVASKTENENIN-----QIGHQEKTKSS 1146
Db 400 BPEAEKKEEDTTEKKSKVGESEEGKGGKKKKDKKNNKKKDKTKPRKMTDEDEBKDD 459

QY 1147 SEENVRS 1154
Db 460 KDVKIEGS 467

RESULT 15
T06310
hypothetical protein F11C18.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 10:14:27 ; Search time 24.77 Seconds
(without alignments)
1680.275 Million cell updates/sec

Title: US-09-775-181-2

Perfect score: 6382

Sequence: 1 MGAMAYPLLCILLLAQLGLG.....LSANKIAPRKEEIWDSFKV 1215

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	220.5	3.5	1199	1 MGR1_RAT	P23385 rattus norv
2	211	3.3	1203	1 MGR5_RAT	P31424 rattus norv
3	199	3.1	704	1 NP14_RAT	P41777 rattus norv
4	197.5	3.1	1212	1 MGR5_HUMAN	P41594 homo sapien
5	193	3.0	1194	1 MGR1_HUMAN	Q13255 homo sapien
6	184	2.9	976	1 MGR_DROME	P91685 drosophila
7	181	2.8	1358	1 SIR4_YEAST	P11978 saccharomyc
8	178	2.8	1637	1 MRSP_STAAU	P60544 staphylococ
9	177.5	2.8	728	1 TRDN_HUMAN	Q13061 homo sapien
10	177.5	2.8	3924	1 ANK2_HUMAN	Q01484 homo sapien
11	177	2.8	667	1 CYL1_BOVIN	P35662 bos taurus
12	176	2.8	2349	1 TPR_HUMAN	P12270 homo sapien
13	175.5	2.7	1079	1 CASR_MOUSE	Q9gy96 mus musculu
14	174.5	2.7	1131	1 AC15_MOUSE	P35601 mus musculu
15	174	2.7	1020	1 NFH_HUMAN	P12036 homo sapien
16	172.5	2.7	633	1 MLH1_TETRA	P40631 tetrahymena
17	172.5	2.7	705	1 TRDN_RABIT	Q28820 oryctolagus
18	172.5	2.7	1079	1 CASR_RAT	P48442 rattus norv
19	169.5	2.7	598	1 NEM1_HUMAN	P35663 homo sapien
20	169.5	2.7	915	1 NEM1_HUMAN	P07197 homo sapien
21	169	2.6	999	1 MGR1_CAEL	Q09630 caenorhabdi
22	168.5	2.6	1233	1 YF16_YEAST	P23597 saccharomyc
23	168.5	2.6	1257	1 RBH1_HUMAN	P29374 homo sapien
24	168.5	2.6	1849	1 IGA4_HAEIN	P45386 haemophilus
25	168	2.6	1411	1 TCOF_HUMAN	Q13428 homo sapien
26	167.5	2.6	816	1 YG3A_YEAST	P53278 saccharomyc
27	167	2.6	836	1 NOT3_YEAST	P06102 saccharomyc
28	166	2.6	1523	1 SON_HUMAN	P18583 homo sapien
29	166	2.6	2375	1 ATRX_HUMAN	P46100 homo sapien
30	165.5	2.6	952	1 IF41_YEAST	P39935 saccharomyc
31	165	2.6	1275	1 TRP_DROME	P19334 drosophila
32	164.5	2.6	589	1 HIBN_XENLA	P06180 xenopus lae
33	164	2.6	2476	1 ATRX_MOUSE	Q61687 mus musculu

ALIGNMENTS

RESULT 1

ID	MGR1_RAT	STANDARD;	PRT;	1199 AA.
AC	P23385;			
DT	01-NOV-1991 (Rel. 20, Created)			
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	METABOTROPIC GLUTAMATE RECEPTOR 1 PRECURSOR.			
GN	GRM1 OR MGLUR1.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;			
RX	MEDLINE=91156047; PubMed=1847995;			
RA	Masu M., Tanabe Y., Tsuchida K., Shigemoto R., Nakanishi S.;			
RT	"Sequence and expression of a metabotropic glutamate receptor.";			
RL	Nature 349:760-765(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	TISSUE=Brain;			
RX	MEDLINE=92022526; PubMed=1656524;			
RA	Houamed K.M., Kuipjer J.L., Gilbert T.L., Haldeman B.A., O'Hara P.J.,			
RA	Mulvihill E.R., Almers W., Hagen F.S.;			
RT	"Cloning, expression, and gene structure of a G protein-coupled			
RT	glutamate receptor from rat brain.";			
RL	Science 252:1318-1321(1991).			
RN	[3]			
RP	ALTERNATIVE SPLICING (ISOFORM 1B).			
RC	TISSUE=Brain;			
RX	MEDLINE=92110002; PubMed=1309649;			
RA	Tanabe Y., Masu M., Ishii T., Shigemoto R., Nakanishi S.;			
RT	"A family of metabotropic glutamate receptors.";			
RL	Neuron 8:169-179(1992).			
RN	[4]			
RP	ALTERNATIVE SPLICING (ISOFORM 1C).			
RC	TISSUE=Brain;			
RX	MEDLINE=93066232; PubMed=1438218;			
RA	Pin J.-P., Waeber C., Prezeau L., Bockaert J., Heinemann S.F.;			
RT	"Alternative splicing generates metabotropic glutamate receptors			
RT	inducing different patterns of calcium release in Xenopus oocytes.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 89:10331-10335(1992).			
RN	[5]			
RP	FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS			
CC	MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-			
CC	CALCIUM SECOND MESSENGER SYSTEM. MAY PARTICIPATE IN THE CENTRAL			
CC	ACTION OF GLUTAMATE IN THE CNS. SUCH AS LONG-TERM POTENTIATION IN			
CC	THE HIPPOCAMPUS AND LONG-TERM DEPRESSION IN THE CEREBELLUM.			
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.			
CC	-!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1A (SHOWN HERE), 1B AND 1C;			
CC	ARE PRODUCED BY ALTERNATIVE SPLICING. 1B AND 1C ARE C-TERMINALLY			
CC	TRUNCATED FORMS OF 1A.			
CC	-!- TISSUE SPECIFICITY: IS PREDOMINANTLY EXPRESSED IN CEREBELLAR			
CC	PURKINJE CELLS, CA2-CA3 PYRAMIDAL CELLS OF THE HIPPOCAMPUS, AND			
CC	MITRAL AND TUFTED CELLS OF THE OLFACTORY BULB.			

Q14832 homo sapien
P46012 caenorhabdi
P31422 rattus norv
Q02952 homo sapien
Q43493 homo sapien
P98193 rattus norv
P30414 homo sapien
P51825 homo sapien
P25386 saccharomyc
P34487 caenorhabdi
Q12263 saccharomyc
Q03661 saccharomyc

CC -!- MISCELLANEOUS: ACTIVATED BY QUISOVALATE > GLUTAMATE > IBOTENATE >
CC TRANS-1- AMINOCYCLOPENTYL-1,3-DICARBOXYLATE; INHIBITED BY
CC 2-AMINO- 3-PHOSPHONOPROPIONATE.
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC STRONGEST, TO MGLUR5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X57569; CAA40799.1; -;
CC EMBL; M61099; AAB19497.1; -;
CC EMBL; S48085; AAB24138.1; -;
CC PIR; S15362; S15362.
CC PIR; A41939; A41939.
CC GCRDB; GCR_0216; -;
CC InterPro; IPR000162; -;
CC InterPro; IPR000337; -;
CC InterPro; IPR001256; -;
CC InterPro; IPR001828; -;
CC Pfam; PF00003; 7tm_3; 1.
CC Pfam; PF01094; ANF_receptor; 1.
CC PRINTS; PR00248; GPCRMR.
CC PRINTS; PR00593; MTABOTROPICR.
CC PRINTS; PR01051; MTABOTROPICR.
CC PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1.
CC PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1.
CC PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
CC PROSITE; PS02059; G_PROTEIN_RECEP_F3_4; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
CC Multigene family; Alternative splicing.
CC SIGNAL 1 18
CC CHAIN 19 1199 METABOTROPIC GLUTAMATE RECEPTOR 1.
CC DOMAIN 19 592 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 593 615 I (POTENTIAL).
CC DOMAIN 616 629 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 630 650 II (POTENTIAL).
CC DOMAIN 651 661 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 662 680 III (POTENTIAL).
CC DOMAIN 681 706 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 707 727 IV (POTENTIAL).
CC DOMAIN 728 750 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 751 772 V (POTENTIAL).
CC DOMAIN 773 785 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 786 808 VI (POTENTIAL).
CC DOMAIN 809 814 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 815 840 VII (POTENTIAL).
CC DOMAIN 841 1199 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 1014 1034 GLN/PRO-RICH.
CC DOMAIN 1074 1080 GLN/PRO-RICH.
CC DOMAIN 1126 1135 ASP/GLU-RICH (ACIDIC).
CC DOMAIN 1140 1199 SER-RICH.
CC CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 223 223 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 397 397 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).
CC VARSPLIC 887 906 NSNGKSVSWSEPGGRQAPK -> KKQPEFSPSQCSAH
CC AQL (IN ISOFORM 1B).
CC VARSPLIC 907 1199 MISSING (IN ISOFORM 1B).
CC VARSPLIC 888 897 SNKGSVSWSE -> FALDRONTYV (IN ISOFORM 1C).
CC VARSPLIC 898 1199 MISSING (IN ISOFORM 1C).
CC SEQUENCE 1199 AA; 133235 MW; EEE5A04C50694B9F CRC64;

Query Match 3.5%; Score 220.5; DB 1; Length 1199;
Best Local Similarity 18.6%; Pred. No. 0.0004;
Matches 163; Conservative 120; Mismatches 330; Indels 265; Gaps 29;
Qy 357 GFYHPGVLPVNNFRRRGPDQHSIGSTKDVSEE-----AYVCLPREG- 398

Db 498 GTWHEGVNIDDDYKIQ-----MNRSGMVRVSCBPCCLKGQIKVRKGEVSCCWTACKENE 554
Qy 399 -----CPFC-----ADDSPCFQVEDKYL-----LAISFGQICMLDLDFVSMLV 438
Db 555 FVQDEFTCRADLGMWPAELTGCEPIPVRYLEWSDIESIIAIAFSCILGILVTLFVTLIF 614
Qy 439 YHFRKAKSIRASGLILLETILFGLSLLLVFPVILYFESTERCILLRWALLGFATVYGT 498
Db 615 VLYRDPVVKSSRELCAIILAGIFLGVCFVTLIAKTITTCYQLORLLVGLSSAMCYSA 674
Qy 499 VTLKLHVRVKV-----FLSRTAQRIPYMTGGRVMRMLAVILLVFWFELGW 544
Db 675 LVTKTNRIARILAGSKKICKTRPFMSAWAQVIAIISLVQLTLVVTLLII----- 726
Qy 545 TSSVCQNLKLEKISLIGQKTSDDLIFNNCLIDRWDMYMTAVA-EFLFLLWGVLYCYAVRTV 603
Db 727 -----MEPPMILTSLYSPIKE--VYLICITNLGVAVPGVYNGLLINSCTYYAFKTRV 777
Qy 604 PSATHEPRYMAVAVHNELIISAIHITIRFVLASRLQSDWMLMLYFAHTHLTVTITGLLL 663
Db 778 PANENEAKYIAFTWYTCIIWLAFVIYF-----GSNYKIITTCFAVLSVYVALGCMF 831
Qy 664 IPK-----FSHSSNNPRDDIATEAYEDELDMG-----RSGSYLNSSINSAMSEHSLDP 711
Db 832 TPKMYIIIAKPERNVSFAFTT---SDVVRMHVGDGKLPKCRSNTFLN----- 874
Qy 712 EDIRDELKLLYAOLEIYKRRKMTNNPHLQKRCS-----KGLGRSIRRRITEIPET 764
Db 875 -----IFRRKKPGAGNANSKSVSWSEPGGRQAPKQGHVWQRLSVHVYKT 919
Qy 765 VSROCKEDKEGADHGTAKGTALIRKNPESNGTSGKKEETLKNRVFSLKSHSTYDHY 824
Db 920 NETACNQ-----TAVIK---PLTKSYQGSCKSLTFSD----- 948
Qy 825 RDQTESSSLPESQEEETTENSTLESLSGKLTQKLEDSEAESESTESVPLVCKSASAHN 884
Db 949 -----ASTKLYNVEEDNTPSAHFSPSPSPSVVHR 980
Qy 885 LSSEKKTCHPRTSMLOKSLSVIASAKETLGLAGTKTAGVEERTKSKOKPLPKOKETNRN 944
Db 981 RGPVATTPPLPHL-----TAETPLFLADSVTPKGL-----PPLPQOO----- 1021
Qy 945 HNSDNTETKDPAPQNSNPAEPRKPKQKIMKQORVNTTANSDL-----NPGTTQ 996
Db 1022 -----PQQPPQPPQPPK--SLMDQLQGVVTFNGSGIPDFHAVLAGPCTP- 1065
Qy 997 MKDNFDIGEVCPWEVYDLTPGPVPSE--SKVKHVSIVASEMEKNPTFSLEKSHHKPKA 1054
Db 1066 -----GN-----SLRSLYPPPPPPQHLMPLHLSTFQEEISIPPGEDIDDD- 1108
Qy 1055 AEVCOQSNQKIDKAEVCLWESQ-----QSILEDEKLLISKTPTVLPRAKEENGQOPRAAN 1111
Db 1109 -----ERFKLLQEFVREGNTEDELEEEEDLPTAKLTPEDSPALTPPSPFRDS 1159
Qy 1112 VCAGQSEELPP--KAVASKTENENLNOIGHOEKKTSSS 1147
Db 1160 VASGSSVSPSPSVESVLCPTPPNVTYASVILRDYKQSSS 1197
RESULT 2
MGR5_RAT
ID MGR5_RAT STANDARD; PRT; 1203 AA.
AC P31424;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 5 PRECURSOR.
GN GRM5 OR MGLUR5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;

[1]
SEQUENCE FROM N.A.
TISSUE=Brain;
MEDLINE=92317054; PubMed=1320017;
Abe T., Sugihara H., Nawa H., Shigemoto R., Mizuno N., Nakanishi S.;
"Molecular characterization of a novel metabotropic glutamate
receptor mGluR5 coupled to inositol phosphate/Ca2+ signal
transduction.";
J. Biol. Chem. 267:13361-13368(1992).
[2]
SEQUENCE OF 859-923 FROM N.A., AND ALTERNATIVE SPLICING.
TISSUE=Brain;
MEDLINE=93343913; PubMed=7688218;
Minakami R., Katsuki F., Sugiyama H.;
"A variant of metabotropic glutamate receptor subtype 5: an
evolutionally conserved insertion with no termination codon.";
Biochem. Biophys. Res. Commun. 194:622-627(1993).
-1- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS
MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-
CALCIUM SECOND MESSENGER SYSTEM AND GENERATES A CALCIUM-ACTIVATED
CHLORIDE CURRENT.
-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 5A (SHOWN HERE) AND 5B; ARE
PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER BY AN INSERTION OF
32 RESIDUES.
-1- TISSUE SPECIFICITY: WIDELY DISTRIBUTED IN NEURONAL CELLS OF THE
CENTRAL NERVOUS SYSTEM.
-1- MISCELLANEOUS: ACTIVATED BY OUTSQUALATE > GLUTAMATE > IBOTENATE >
TRANS-1- AMINOCYCLOPENTYL-1,3-DICARBOXYLATE.
-1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
STRONGEST, TO MGLUR1.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL; D10891; -; NOT_ANNOTATED_CDS.
DR EMBL; S64315; AAB27666.1; -;
DR PIR; A42916; A42916.
DR GCRDB; GCR_0444; -;
DR GCRDB; GCR_0760; -;
InterPro: IPR000162; -;
InterPro: IPR000202; -;
InterPro: IPR000337; -;
InterPro: IPR001828; -;
Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMR.
DR PRINTS; PR00593; MTABOTROPICR.
DR PRINTS; PR01055; MTABOTROPIC5R.
DR PROSITE; PS00979; G_PROTEIN_RECP_F3_1; 1.
DR PROSITE; PS00980; G_PROTEIN_RECP_F3_2; 1.
DR PROSITE; PS00981; G_PROTEIN_RECP_F3_3; 1.
DR PROSITE; PS00981; G_PROTEIN_RECP_F3_3; 1.
DR PROSITE; PS00981; G_PROTEIN_RECP_F3_4; 1.
G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
Multigene family; Alternative splicing.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1203 METABOTROPIC GLUTAMATE RECEPTOR 5.
FT DOMAIN 22 578 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 579 601 I (POTENTIAL).
FT DOMAIN 602 615 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 616 636 II (POTENTIAL).
FT DOMAIN 637 647 III (POTENTIAL).
FT TRANSMEM 648 666 IV (POTENTIAL).
FT DOMAIN 667 692 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 693 713 V (POTENTIAL).
FT DOMAIN 714 736 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 737 758 V (POTENTIAL).
FT DOMAIN 759 771 CYTOPLASMIC (POTENTIAL).

FT	TRANSMEM	772	794	VI (POTENTIAL).
FT	DOMAIN	795	800	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	801	826	VII (POTENTIAL).
FT	DOMAIN	827	1203	CYTOPLASMIC (POTENTIAL).
FT	CARBOHYD	88	88	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	209	209	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	377	377	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	381	381	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	444	444	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	733	733	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPIC	876	907	MISSING (IN ISOFORM 5A).
SQ	SEQUENCE	1203	AA; 131885	MW; 99CA51E9E11C1E44 CRC64;

Query Match 3.3%; Score 211; DB 1; Length 1203;
Best Local Similarity 18.7%; Pred. No. 0.0012;
Matches 179; Conservative 128; Mismatches 400; Indels 248; Gaps 36;

QY	219	ETEFHGLRRKWRPHLHRRGPNQGRGLGHSWR-----RKDGLGDKSHKFWPPYLECE	273
DB	332	DVKWFDDYYLKLRLPETNLRNP-----WFQEFQHRFQCRLEGFAQENSKYNT-----CN	381
QY	274	NG-----SYKPGWLIV--TLSSAIYGL--OPNLVPEFRGV-----MKVDINLQ	311
DB	382	SSLTLRTHVQDSKMGFEVINAIYTMAYGLHNMQMSLCPCYAGLGDAMKPDIGRKLDSLM	441
QY	312	KVDIDQSSDGFSGTHKCHLNSECMPKGLGFLVGLGAYECICKAGFYHPGVLPV-----	366
DB	442	KTNFTGVGDMILFDENGSDSPGRYEIMNFKEG--KDYFDYINVGSMDNGELKMDDEV	498
QY	367	----NPNRRGPDQHI--SGSTKDVSEE-----AVCLPCREG-----CPFC-----A	403
DB	499	WSKKNNTIIRSCPECKGQIKVIRKEVSCCWCTCTCKENYEVDFEYTCACOLGSWPT	558
QY	404	DD--SPCFVQEDKYLR-----LAIISFGICMLLDFVSMVYVHFRKAKSIRASGLILLE	456
DB	559	DDLTGCDLIPVQLRGDPPIAAVFPACUGLLATLEVTVIIFIRTPVPPVSSSRRELCY	618
QY	457	TLIFGSLLYFPVVIIVFEPSTFCILLRWARLIGFATVYGTVTIKLHRLKV-----	509
DB	619	IILAGICLGYLCTFCLIAKPKQIYCYLORIGLSPAMSYSALVTKTNRILARILAGSKK	678
QY	510	-----FLSRTAQRIPTYTGGVRMRMLAVILLVWFVLGWTSSVCONLEKQISL-----	558
DB	679	ICTKKPRFMSACAQ-----LVIAPIIL-----CIQLGIIVALTME	714
QY	559	---IGOKTSDHLIFNNCLIDRWDTAVA-EFLFLMGVYLCYAVRTVPSAFHEPYMA	614
DB	715	PPDLMHDYPSIREVYLICNTNLGVVTPGLYNGLLILISCTFYAFKTRNVFANFNEAKYIA	774
QY	615	VAVHNELIISAIFHTIRFVLASRLQSDWMLMYFAHTVTLTVTITGLLLIPK-----FSHS	670
DB	775	FTMYTTCIIWLAFPVIYF-----GSNYKIITWCFVSLSATVALGCMFPVKVYIILAKP	828
QY	671	SNPNRDDIATE-AYEDELDMGRSGSYL--NSSINSANSEHSLEDIEDIRLEKIL---YAQ	724
DB	829	ERNVRSAPFTTSTVVRMHVGDGKSSAASRSSSLVNLKRRGSSGETLRYKDRRLAOKSE	888
QY	725	LEIYKRRKMTNPNHPLQKCKSKKGL-----GRSIRMRRIT-----EIP-ETVSR	767
DB	889	IECTPKSGMNGRATMSSSSNGKSVTWAOANEKSTRQHLWRLSVUINKENPNQAVI	948
QY	768	QCSKEDEKADHGTAKG-----TALIRKNPPESSGNTGKSKEETLKNRVF	812
DB	949	KPFKSTENRGPCAAAGGSGPGVAGAGNAGCTATGGPEPPDAGPKA-----	995
QY	813	SLKKSHTYDHRDQDTSESSSLPTESQEEETENSTLESGLKKLTKLKEDSSESTES	872
DB	996	-----LYDVAEAEESFPAAARPRSPSPITSLHLAG-----SAGRTDDDD	1034
QY	873	VPLVCKSASAHNLSSEKKTGHPRTSMLOKLSIVTASAKEKTLGLAG-----	918
DB	1035	APSLHSETAARSSSSQGLMEQISSVVVTRFTANISLNMMLSTAAATPGPGTICSSYL	1094

QY 919 -----KTQAGVEERTKSKQPLPKDKETNRNHSNDNTETKDPAPNSNPAAE-PRK 969
Db 1095 IPKEIQLEPTTMTTFAE-----IQPLPAIEVT-----GGAQAGTGVSPAQETPTGAESAPCK 1145
QY 970 POKSGIMKOORVNPITANSNDLNGTQMDNFDCVCPWEVYDITPGVPVPSK 1024
Db 1146 PDLEELVALTPPSP-----FRDSDVSGSTTTPNSPVSEALCIPSPK 1187

RESULT 3
NP14_RAT

ID NP14_RAT STANDARD: PRT: 704 AA.
AC P41777;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NUCLEOLAR PHOSPHOPROTEIN P130 (NUCLEOLAR 130 KDA PROTEIN) (140 KDA
DE NUCLEOLAR PHOSPHOPROTEIN) (NOPP140).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
NCBI_TaxID=10116;
[1]

SEQUENCE FROM N.A., AND SEQUENCE OF 292-309 AND 563-601.

RC TISSUE=Liver;
RX MEDLINE=92323542; PubMed=1623516;
RA Meier U.T., Blobel G.;
RT "Nopp140 shuttles on tracks between nucleolus and cytoplasm.";
RL Cell 70:127-138(1992).
CC -!- FUNCTION: RELATED TO NUCLEOGENESIS, MAY PLAY A ROLE IN THE
CC MAINTENANCE OF THE FUNDAMENTAL STRUCTURE OF THE FIBRILLAR CENTER
CC AND DENSE FIBRILLAR COMPONENT IN THE NUCLEOLUS. IT HAS INTRINSIC
CC GTPASE AND ATPASE ACTIVITIES. MAY PLAY AN IMPORTANT ROLE IN
CC TRANSCRIPTION CATALYZED BY RNA POLYMERASE I (BY SIMILARITY).
CC -!- SUBUNIT: INTERACTS WITH DKC1/NAP57.
CC -!- SUBCELLULAR LOCATION: SHUTTLES ON CURVILINEAR TRACKS BETWEEN
CC NUCLEOLUS AND CYTOPLASM. THESE TRACKS EXTEND FROM THE DENSE
CC FIBRILLAR COMPONENT OF THE NUCLEOLUS ACROSS THE NUCLEOLAR PLASM TO
CC A LIMITED NUMBER OF NUCLEAR PORE COMPLEXES.
CC -!- PTM: THIS PROTEIN UNDERGOES RAPID AND MASSIVE PHOSPHORYLATION AND
CC DEPHOSPHORYLATION ON CK-II AND PKC SITES. NOPP140 IS ONE OF THE
CC MOST PHOSPHORYLATED PROTEINS IN THE CELL.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M94287; AAA41718.1; -;
DR EMBL; M94288; AAA41719.1; -;
KW Nuclear protein; Phosphorylation; Repeat; GTP-binding; ATP-binding.
FT DOMAIN 84 570
FT 11 X 12 AA APPROXIMATE REPEATS OF AN

FT REPEAT 84 95
FT ACIDIC SERINE CLUSTER 1.
FT REPEAT 127 138
FT ACIDIC SERINE CLUSTER 2.
FT REPEAT 170 181
FT ACIDIC SERINE CLUSTER 3.
FT REPEAT 231 242
FT ACIDIC SERINE CLUSTER 4.
FT REPEAT 274 285
FT ACIDIC SERINE CLUSTER 5.
FT REPEAT 335 346
FT ACIDIC SERINE CLUSTER 6.
FT REPEAT 373 384
FT ACIDIC SERINE CLUSTER 7.
FT REPEAT 434 445
FT ACIDIC SERINE CLUSTER 8.
FT REPEAT 479 490
FT ACIDIC SERINE CLUSTER 9.
FT REPEAT 524 535
FT ACIDIC SERINE CLUSTER 10.
FT REPEAT 559 570
FT ACIDIC SERINE CLUSTER 11.
FT MOD_RES 567 567
FT PHOSPHORYLATION (BY CK2).
FT VARIANT 150 150
FT MISSING (IN NOPP140B).
SQ SEQUENCE 704 AA; 73562 MW; 14DF1BF2DE4B3EA3 CRC64;

Query Match 3.1%; Score 199; DB 1; Length 704;
Best Local Similarity 21.3%; Pred. No. 0.0026;
Matches 126; Conservative 92; Mismatches 240; Indels 134; Gaps 24;
QY 665 PKFSHSSNNP-----RDIATEAYED---ELDMGR-----SGSYLN 597
Db 64 PKVKLOSNGPVAKKAKKETSSDSSDSEEDKAQVPTQKAAAPAKRASLPQHGAKAAA 123
QY 698 SSINSAMSEHSLDPEDIRDELKLYAQLEIYKRRKMTNNPHLOKKGKSKGLGHSIMRR 757
Db 124 KASESSSESESEEEEEKKKKPVQOKAVKPOAKAVRPPP---KKAESSESSESSSE- 179
QY 738 ITEIPETVSROCKEDKEGADHGTAKGTALIRKNPESPESGNTGSKKEETLKNRVSLKKS 817
Db 180 -DEAPQT-----QKPKAAATAAKAPTQAKTAPAKPGPPAKPAKAAANGKAGSSSS 230
QY 818 HSTYDHRVDQTEESSLSLTESQEBEETTENTSTLESLSGKK-LTOKLKEDSEARSTESVPLV 876
Db 231 -----SSSSSSDSEEEKKAAAPLKKTPAKKKVYVAKAPVKVTAAPTQ----- 272
QY 877 CKSASAHNLSE-----KKTGHPTSMLOKSLSVIASAKEKTLGLAGKTQTAGVEE 927
Db 273 -KSSSESSSESEEEEEQKKPMKKKAGPYSSVPPPSVSL-----SKKSVG----- 315
QY 928 RTKSQKPLPKDKETNRNHSNDNTETKDPAPNSNPAAEPRKPKQSGIMK-QOORVNPPTA 986
Db 316 ---AQSPKAAATQPADSSADSSSESD-----SSSEEEKTKTAKTVVSKTAKPAPVKK 367
QY 987 NSDLNPGTITQMDNFDCVCPWEVYDITPGVPVPSKVKQHVSIIVASEMEKNPTFSIKE 1046
Db 368 KAE-----SSSDSSDSSDEAPAKPVSAKPSLSPKPAVTPKPPAAKAVATPKQPA----- 419
QY 1047 KSHHKPAAEVCQSNOKRIDKAEVCLWESOGOSTLEDEKLLISK-TPVLPRAKEENG 1105
Db 420 GSGQKQPKKADSSSESESSSESE---EATKSKVTTPKARVTAATAAASLPAKQAPRAGG 476
QY 1106 QPRAANVCAGOSE--LPPKAVASK-----TENENLNQIGHQEKTKTSSSEENVRGSYN 1156
Db 477 DSSSDSSSESESEEEKTKPPKPAKKAAGAAVPKTPVKKAAAESESSSSSE-----D 530
QY 1157 SSNFOQPLTSRAEVCVPWEFETPAQPNAGRSVALPASPALSANKIAGPRKEE 1208
Db 531 SSEEKKKPKSKA-----TP-KPQAGKANGVPASQ-----NGKAGKESEE 569

RESULT 4

MGR5_HUMAN
ID MGR5_HUMAN STANDARD: PRT: 1212 AA.
AC P41594;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 5 PRECURSOR.
GN GRM5 OR MGLUR5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94197696; PubMed=7908515;
RA Minakami R., Katsuki F., Yamamoto T., Nakamura K., Sugiyama H.;
RT "Molecular cloning and the functional expression of two isoforms of
RL human metabotropic glutamate receptor subtype 5";
RL Biochem. Biophys. Res. Commun. 199;1136-1143(1994).
[2]
RN REVISONS..
RP Katsuki F.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 860-952 FROM N.A.
RC TISSUE=Brain;

MEDLINE=93343913; PubMed=76982218;
Minakami R., Katsuki F., Sugiyama H.;
"A variant of metabotropic glutamate receptor subtype 5: an evolutionally conserved insertion with no termination codon."; Biochem. Biophys. Res. Commun. 194:622-627(1993).
-!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM AND GENERATES A CALCIUM-ACTIVATED CHLORIDE CURRENT.
-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
-!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 5A (SHOWN HERE) AND 5B; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER BY AN INSERTION OF 32 RESIDUES.
-!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS. STRONGEST, TO MGLUR1.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; D28538; BAA05891.1; -;
EMBL; D28539; BAA05892.1; -;
EMBL; S84316; AADI3954.1; -;
GCRDB; GCR_0761; -;
GCRDB; GCR_1002; -;
GCRDB; GCR_1003; -;
GCRDB; GCR_1317; -;
MIM; 604102; -;
InterPro; IPR000162; -;
InterPro; IPR000202; -;
InterPro; IPR000337; -;
InterPro; IPR001828; -;
Pfam; PF00003; 7tm_3; 1;
Pfam; PF01094; ANF_receptor; 1;
PRINTS; PR00248; GPCRNGR.
PRINTS; PR00593; MTABOTROPICR.
PRINTS; PR01055; MTABOTROPICSR.
PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1;
PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1;
PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1;
PROSITE; PS0259; G_PROTEIN_RECEP_F3_4; 1;
G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
Multigene family; Alternative splicing.
CHAIN 1 20 POTENTIAL.
FT CHAIN 21 1212 METABOTROPIC GLUTAMATE RECEPTOR 5.
FT FT DOMAIN 22 579 EXTRACELLULAR (POTENTIAL).
FT FT TRANSMEM 580 602 I (POTENTIAL).
FT FT DOMAIN 603 616 CYTOPLASMIC (POTENTIAL).
FT FT TRANSMEM 617 637 II (POTENTIAL).
FT FT DOMAIN 638 648 EXTRACELLULAR (POTENTIAL).
FT FT TRANSMEM 649 667 III (POTENTIAL).
FT FT DOMAIN 668 693 CYTOPLASMIC (POTENTIAL).
FT FT TRANSMEM 694 714 IV (POTENTIAL).
FT FT DOMAIN 715 737 EXTRACELLULAR (POTENTIAL).
FT FT TRANSMEM 738 759 V (POTENTIAL).
FT FT DOMAIN 760 772 CYTOPLASMIC (POTENTIAL).
FT FT TRANSMEM 773 795 VI (POTENTIAL).
FT FT DOMAIN 796 801 EXTRACELLULAR (POTENTIAL).
FT FT TRANSMEM 802 827 VII (POTENTIAL).
FT FT DOMAIN 828 1212 CYTOPLASMIC (POTENTIAL).
FT FT CARBOHYD 88 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 210 210 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 378 378 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 382 382 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 445 445 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 734 734 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 877 908 MISSING (IN ISOFORM 5A).
SQ SEQUENCE 1212 AA; 132468 MW; A3C7360681C6A25 CRC64;

QY 1106 QPRAANVCAG-----QSEELPPKAVASKTENENLQIGHOEKKTSSSEENVVRCYSNNNF 1161
DB 1128 OPAQAQAQAADARESPAAGPAAAKPDLELVAL-----TPSPFRSDVDSGGST- 1178
QY 1162 QOPLTSRAEVCWPEFTPAQPNAGRSVALPAS 1193
DB 1179 -----TPNSPVSEALCIPSS 1194

RESULT 5
MGR1_HUMAN
ID MGR1_HUMAN STANDARD: PRT: 1194 AA.
AC Q13255: Q13256; Q14757; Q14758;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 1 PRECURSOR.
GN GRM1 OR MGLR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. PubMed=7476890;
RA MEDLINE=96029774; PubMed=7476890;
RA Desai M.A., Burnett J.P., Wayne N.G., Schoepp D.D.;
RT "Cloning and expression of a human metabotropic glutamate receptor 1
RT alpha: enhanced coupling on co-transfection with a glutamate
RT transporter";
RL Mol. Pharmacol. 48:648-657(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97231349; PubMed=9076744;
RA Stephan D., Bon C., Holzwarth J.A., Galvan M., Pruss R.M.;
RT "Human metabotropic glutamate receptor 1: mRNA distribution,
RT chromosome localization and functional expression of two splice
RT variants";
RL Neuropharmacology 35:1649-1660(1996).
CC -!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS
CC MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-
CC CALCIUM SECOND MESSENGER SYSTEM. MAY PARTICIPATE IN THE CENTRAL
CC ACTION OF GLUTAMATE IN THE CNS, SUCH AS LONG-TERM POTENTIATION IN
CC THE HIPPOCAMPUS AND LONG-TERM DEPRESSION IN THE CEREBELLUM.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC STRONGEST, TO MGLR5.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: U31215; AAA87843.1; -
CC EMBL: U31216; AAA87844.1; -
CC EMBL: L76627; AAB05337.1; -
CC EMBL: L76631; AAB05338.1; -
CC MIN: 604473; -
CC GCRDb; GCR1825; -
CC GCRDb; GCR1826; -
CC GCRDb; GCR1982; -
CC GCRDb; GCR1983; -
CC InterPro; IPR000162; -
CC InterPro; IPR000337; -
CC InterPro; IPR001256; -
CC InterPro; IPR001828; -
CC Pfam; PF00003; 7tm.3; 1.
CC Pfam; PF01094; ANF_receptor; 1.
CC PRINTS; PR00248; GPCRMR.

DR PRINTS; PR00593; MTABOTROPICR.
DR PRINTS; PR01051; MTABOTROPICR.
DR PROSITE; PS00979; G-PROTEIN_RECEP_F3_1; 1.
DR PROSITE; PS00980; G-PROTEIN_RECEP_F3_2; 1.
DR PROSITE; PS00981; G-PROTEIN_RECEP_F3_3; 1.
DR PROSITE; PS00982; G-PROTEIN_RECEP_F3_4; 1.
DR PROSITE; PS00983; G-PROTEIN_RECEP_F3_5; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Multigene family; Alternative splicing.
FT SIGNAL 1 18
FT CHAIN 19 1194 METABOTROPIC GLUTAMATE RECEPTOR 1.
FT DOMAIN 19 592 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 593 615 I (POTENTIAL).
FT DOMAIN 616 629 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 630 650 II (POTENTIAL).
FT DOMAIN 651 661 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 662 680 III (POTENTIAL).
FT DOMAIN 681 706 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 707 727 IV (POTENTIAL).
FT DOMAIN 728 750 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 751 772 V (POTENTIAL).
FT DOMAIN 773 785 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 786 808 VI (POTENTIAL).
FT DOMAIN 809 814 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 815 840 VII (POTENTIAL).
FT DOMAIN 841 1194 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1014 1035 GLN/PRO-RICH.
FT DOMAIN 1067 1081 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 1095 1130 SER-RICH.
FT DOMAIN 1142 1194 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 223 223 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 397 397 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPPLIC 907 906 NSNGKSVSWSPGGGVKPK -> KKQPEFSPSQCSAH
FT VARSPPLIC 887 906 AQL (IN ISOFORM BETA).
FT VARSPPLIC 887 906 MISSING (IN ISOFORM BETA).
FT VARSPPLIC 887 906 NSNGKSVSWSPGGGVKPK -> KKQPEFSPSQCSAH
FT CONFLICT 593 593 VQL (IN ISOFORM BETA FROM REF. 2).
FT SEQUENCE 1194 AA; 132376 MW; 970551AF40584F40 CRC64;

Query Match 3.0%; Score 193; DB 1; Length 1194;
Best Local Similarity 18.7%; Pred. No. 0.01;
Matches 165; Conservative 114; Mismatches 326; Indels 276; Gaps 29;
QY 357 GFYHPGVLVNNFRERRGPDQHSSTKDVSE-----AYVCLPCREG- 398
DB 498 GTWEGVLNDDYKIQ---MNKSGVVRVSVCEPCLKQIKVIRKGEVSCCMCTACKENE 554
QY 399 -----ADSPCFVQEDKYLR-----LATISFQGLCMLDFVSMLVV 438
DB 555 YVDEFTCKACDLGWPNADLTGCEPIPVRYLEWSNIEPITAIEFSCIGILVLFVTLIF 614
QY 439 YHFRKAKSIRASGLILLLETLFGSLLYFPVVLVYFEPSTFCIRLLRWARLGFATVYGT 498
DB 615 VLYRDPVVKSSRELVCILLAGIFLGVGVCPETLAKPTTSCYLRLLVGLSSAMCYSA 674
QY 499 VTLKLRHLVKV-----FLSRTAQRIPTMTGGRVWRMIAVILLVFWFELGW 544
DB 675 LVTKTNRIARILAGSKKICKTRKPRFMSAWAQVITIASILSVQLTLVTLII----- 726
QY 545 TSSVCQNLKQISLIGQCKTSDHLIFNCLIDRWDMYMTVAEAF-----LFLMGVYLCVAV 600
DB 727 -----MEPPMIL-----SYPSIKEVYLICNTSNLGVAPLGYNGLIISCTYAEKT 774
QY 601 RTVPSAFHEPRYMAVAVINELIISAIFHTIRFVLASRLQSDMMLLYFAHTHLVTVTIG 660
DB 775 RNVPANENEAKYIAFTMTTCIILWLAFAVPIVF-----GSNYKIITTCFAVSLSVTVALG 828
QY 661 LLLTPK----FSSNSNPRDDIATEAYEDELDMG-----RSGSYLNSINSINSAWSEHS 708
DB 829 CMFTPKMYIIITAKPERNVNSAFTT---SDVVRMHVGDGKLPGRSNTFELN----- 874

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC		EMBL; AF115379; AAD09131.1; -.	
DR		pfam: PF00746; Gram_pos_anchor: 1.	
DR		PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.	
KW		Cell wall; Transmembrane; Antibiotic resistance; Glycoprotein; Repeat;	
KW		Signal.	
FT	SIGNAL	1 48	POTENTIAL.
FT	CHAIN	49 1637	METHICILLIN-RESISTANT SURFACE PROTEIN.
FT	DOMAIN	1301 1582	141 X 2 AA TANDEM REPEATS OF D-[SAG].
FT	DOMAIN	1598 1603	CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT			PROTEINS.
SQ	SEQUENCE	1637 AA; 174573 MW; 75BE9ADB469BD309 CRC64;	

Query Match	2.88;	Score 178;	DB 1;	Length 1637;
Best Local Similarity	20.8%;	Pred. No. 0.093;		
Matches 99;	Conservative	71;	Mismatches 197;	Indels 108;
				Gaps 17;

```

763  ETVSRQCKEDKEGADHGHTAKGTALIRKPNPSSGNTGKSKEETLKNRVFSLKXSHSTYD  822
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
47  EAAENNTTQKDDSDASKVKGVN---QTIEQSSAN--SNESDIPBQV-----D  90
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
823  HVRDOTESSS-----LPTSEQBRETENSTLSLCKKLTOKL-----EDSEA  867
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
91  VTRDTTQASSTEEKANTTEQASSTEEKADTTQATTEAPKAEGTKVETEAPKAETDK  150
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
868  ESTESVPLVCKSAHNLSEKKTGHPRTSMLOKSLVIASAKET-----LGLACKT  920
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
151  ATTEAP---KAEETDKATDEA---PKTEETDKATTEEAPAAEETSKAATEAPKAET  203
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
921  QTAGVEERTKSO-----RPLPKDKETNRNHISN---DNTETKDPAPQNSNPAAEPRKPQ  971
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
204  SKAATEEAPKAEEETKATTEAPKAETEDTKVETEAPKAETSKAATEKAPKAETNKVE  263
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
972  KSGIMKOORVNPNTTANSDLNPQTGMKDNDFIGVCPWCVYDLTPGPVPSPSKVKQHVSI  1031
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
264  TEAPAAEETNKAATEETPAVEDTNAKSN-----SNAQPSETERTQVVDT  308
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1032  VASMEKNPFTSLKEKSH-HKPKAAEFVCOQSNQKRIDKAEVCLWESOGOSTLEDEKLLIS  1090
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
309  VAKDLYKKSEVTEAEKAIEIKVLPKDINSLSNE-----EIKKIALS  349
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1091  KTPVLPRAKEENGOPRAANVCAGQSEELPPKAVASKTENENLNOIGHOEKKTSSSEEN  1150
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
350  E--VLKETANKEN-AQPRATF-----RSVSSNARTTNVYSATALRAAAQDTVT  395
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1151  VRGSYNSNNFPQPLTGRAEVCWPE-----FETPAQPNAGRSVALPASSALSANK  1200
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
396  KKGFGNFTAGHDILHKTYKEEFPNEGTLTAFNTFNENTGTGKGALEYNDKIDFNK  1450

```

```

RESULT 9
TRDN_HUMAN ID TRDN_HUMAN STANDARD; PRT; 728 AA.
AC Q13061;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRIADIN.
GN TRDN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
|||
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=96061957; PubMed=7588753;
RA Taske N.L., Eyre H.J., O'Brien R.O., Sutherland G.R., Denborough M.A.,
RA Foster P.S.;

```

"Molecular cloning of the cDNA encoding human skeletal muscle triadin and its localisation to chromosome 6q22-6q23."; Eur. J. Biochem. 233:258-265(1995).

CC -I- FUNCTION: MAY BE INVOLVED IN ANCHORING CALSEQUESTRIN TO THE JUNCTIONAL SARCOPLASMIC RETICULUM AND ALLOWING ITS FUNCTIONAL COUPLING WITH THE RYANODINE RECEPTOR (BY SIMILARITY).

CC -I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN, SARCOPLASMIC RETICULUM.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

[illegible][illegible]


```
QY 1132 ENLNQIGHQEKTSSEENVGRSYNNFQPLTSRAEYCVWETETPAQPNAGRS---V 1188
Db 610 -----GKKTETSEKESKADMKHLREKVKSTRKESLQHLHNVTKAEKPARVSKDVE 661
QY 1189 ALPAS-SALSANKIAGPKREEIWDSF 1213
Db 662 DVPASKAKEGTEDVSPTKQKSPISF 687

RESULT 10
ANK2_HUMAN
ID ANK2_HUMAN STANDARD; PRT: 3924 AA.
AC Q01484; Q01485;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ANKYRIN 2 (BRAIN ANKYRIN) (ANKYRIN B) (ANKYRIN, NONERYTHROID).
GN ANK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain stem;
RX MEDLINE=94075409; PubMed=8253844;
RA Chan W., Kordeli E., Bennett V.;
RT "440-kD ankyrinB: structure of the major developmentally regulated
RT domain and selective localization in unmyelinated axons.";
RL J. Cell Biol. 123:1463-1473(1993).
[2]
SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RN
RP TISSUE=Brain stem;
RX MEDLINE=91302466; PubMed=1830053;
RA Otto E., Kunimoto M., McLaughlin T., Bennett V.;
RT "Isolation and characterization of cDNAs encoding human brain
RT ankyrins reveal a family of alternatively spliced genes.";
RL J. Cell Biol. 114:241-253(1991).
[3]
REVSIONS.
RA Carpenter S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE OF 463-495 FROM N.A.
RX MEDLINE=92009921; PubMed=1833308;
RA Tse W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,
RA Lux S.E., Ward D.C., Forget B.G.;
RT "Isolation and chromosomal localization of a novel nonerythroid
RT ankyrin gene.";
RL Genomics 10:858-866(1991).
-!- FUNCTION: ANKYRINS ATTACH INTEGRAL MEMBRANE PROTEINS TO CYTO-
SKELETAL ELEMENTS; THEY BIND TO THE ERYTHROCYTE MEMBRANE PROTEIN
BAND 4.2, TO NA-K ATPASE, TO THE LYMPHOCYTE MEMBRANE PROTEIN GP85,
AND TO THE CYTOSKELETAL PROTEINS FODRIN, TUBULIN, VIMENTIN AND
DESMIN. ERYTHROCYTE ANKYRINS ALSO LINK SPECTRIN (BETA CHAIN) TO
THE CYTOPLASMIC DOMAIN OF THE ERYTHROCYTES ANION EXCHANGE PROTEIN;
THEY RETAIN MOST OR ALL OF THESE BINDING FUNCTIONS.
-!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
PRODUCED BY ALTERNATIVE SPLICING.
-!- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
CELLS THROUGHOUT THE BRAIN.
-!- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE
AND FUNCTION (POTENTIAL).
-!- SIMILARITY: CONTAINS 23 ANK REPEATS.
-!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
-----
This SWISS-prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/)
```

or send an email to license@isb-sib.ch).

```
CC
CC EMBL; Z26634; CAB42644.1; -.
DR EMBL; X56957; CAA40278.1; -.
DR EMBL; X56958; CAA40279.2; -.
DR EMBL; M37123; AAA62828.1; -.
DR PIR; S14533; S14533.
DR PIR; A39643; A39643.
DR PIR; B39643; B39643.
DR PIR; S14569; S14569.
DR HSP; Q00420; IAWC.
DR MIM; 106410; -.
DR InterPro; IPR000488; -.
DR InterPro; IPR000906; -.
DR InterPro; IPR002110; -.
DR Pfam; PF00791; ZU5; 1.
DR Pfam; PF00023; ank; 22.
DR Pfam; PF00531; death; 1.
DR PROSITE; PS50088; ANK_REPEAT; 20.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50017; DEATH_DOMAIN; 1.
KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat;
KW Phosphorylation; Multigene family;
FT REPEAT 63 92 ANK 1.
FT REPEAT 96 125 ANK 2.
FT REPEAT 129 158 ANK 3.
FT REPEAT 162 191 ANK 4.
FT REPEAT 193 220 ANK 5.
FT REPEAT 232 261 ANK 6.
FT REPEAT 265 294 ANK 7.
FT REPEAT 298 327 ANK 8.
FT REPEAT 331 360 ANK 9.
FT REPEAT 364 393 ANK 10.
FT REPEAT 397 426 ANK 11.
FT REPEAT 430 459 ANK 12.
FT REPEAT 463 492 ANK 13.
FT REPEAT 496 525 ANK 14.
FT REPEAT 529 558 ANK 15.
FT REPEAT 562 591 ANK 16.
FT REPEAT 595 624 ANK 17.
FT REPEAT 628 657 ANK 18.
FT REPEAT 661 690 ANK 19.
FT REPEAT 694 723 ANK 20.
FT REPEAT 727 756 ANK 21.
FT REPEAT 760 789 ANK 22.
FT REPEAT 793 822 ANK 23.
FT DOMAIN 1773 1950 REPEAT-RICH REGION.
FT REPEAT 1773 1784 REPEAT A.
FT REPEAT 1785 1796 REPEAT A.
FT REPEAT 1797 1808 REPEAT A.
FT REPEAT 1809 1820 REPEAT A.
FT REPEAT 1821 1832 REPEAT A.
FT REPEAT 1833 1844 REPEAT A.
FT REPEAT 1845 1856 REPEAT A.
FT REPEAT 1857 1867 REPEAT A (APPROXIMATE).
FT REPEAT 1868 1879 REPEAT A.
FT REPEAT 1880 1891 REPEAT A (APPROXIMATE).
FT REPEAT 1892 1902 REPEAT A.
FT REPEAT 1903 1914 REPEAT A.
FT REPEAT 1915 1926 REPEAT A.
FT REPEAT 1927 1938 REPEAT A.
FT REPEAT 1939 1950 REPEAT A.
FT DOMAIN 3536 3620 DEATH DOMAIN.
FT VARSPLIC 1039 1039
FT VARSPLIC 1444 3528 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT CONFLICT 475 476 Q -> PE (IN REF. 4).
FT CONFLICT 971 971 I -> S (IN REF. 2).
FT CONFLICT 3581 3582 QY -> HA (IN REF. 2).
FT CONFLICT 3586 3586 I -> Y (IN REF. 2).
SQ SEQUENCE 3924 AA; 430337 MW; 52AC496C428E29D2 CRC64;
```



```
Query Match      2.8%; Score 177.5; DB 1; Length 3924;
Best Local Similarity 21.3%; Pred. No. 0.3;
Matches 123; Conservative 85; Mismatches 230; Indels 139; Gaps 26;

QY 691 RSGSYL--NSSINSAWSEHSL-----DPEDIRDELKLYAQLEI----- 727
Dl 1534 RSGCTRDESSVSSRSEGLVEEWIVSDDELEEARQK--APLEITEYPCVEVRIDKE 1591
QY 728 -----YKRRKMITNPN-----HLQK-----KRCSSKKGLGRSIRMRITPEIPE 763
Dl 1592 IKGKVEKSTGLVNYLDDNTCVPLPKEQLQTVODKAGKKCEALAVGRSSEKGDIPP 1651
QY 764 TVSRQCSKDEKADHCTAGKATLIRKNPPSSNGTCKSKEETLKNRV--FSLKKSSTY 821
Dl 1652 DETOSTOKQHPSL--GIKKP--VRRKLKEQ-----KOREEGLOASAERAKELKGSSE 1702
QY 822 DHVRDQTEESSSLPTSEQEBETTENTSTLESLSGK-KLTKRLKEDSEASTESVPLVCKS- 879
Dl 1703 SLGEDPGLAPEPLTVKATSPLEETPIGSIKDKVKALQKRVED-EQKGRSKLPIRVKKG 1761
QY 880 -----ASAHNLSEKKTGHIPRTSMLOKLSVSIASAKEK---TLGLAGTKTQ-- 921
Dl 1762 EDVPKKTTHRPHAPASPSLKSERHAPGSPKTERHSTLSSAKTERHPPVPSKTEKH 1821
QY 922 -TAGVEERTKSQPLPKDKETNRNHSNDTETKDPAPONSNAEPRKPKQSG-INKQ 979
Dl 1822 SPVSPSAKTERHSPASSSTKESHPSPSTKTERHSPVSTKTERHPPVPSGKTDKRP 1881
QY 980 RVNP---TTANSOLNPTQTKMKNFDIGECVPEVYDLTPG-----PVPSESKVKHV 1029
Dl 1882 PVPSPGRTKHPVSPORTEKR-----LPVPSGRTDKHPVSTAGKTEKHL 1928
QY 1030 SIVAS-EMENKPTFLSLEKSHHKKPKAAEVQC-----QSNQKRIDKAEVCLWE----- 1075
Dl 1929 PVPSPGRTKHPVSPORTEKR-----LPVPSGRTDKHPVSTAGKTEKHL 1928
QY 1076 --SOGQSILEDEKLLISKTPVPERAKEENGQGPRAANVAGQSEELPPRAVASKTENEN 1133
Dl 1988 POBKGVRYEKE-----KGPILTQREAO-----KTENQTKRGQRLPVTGTAEKRGVR 2036
QY 1134 LNOIGHQEKTSSEENVRGSYNNFNQOPLTSRAE 1170
Dl 2037 VSSIGVKKEDAAGGKEV-----LSHKIPERVQSVPE 2068

RESULT 11
CYLL_BOVIN STANDARD; PRT; 667 AA.
AC P35662;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CYLICIN I (MULTIPLE-BAND POLYPEPTIDE I).
GN CYLL OR CYL.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Testis;
RX MEDLINE=93359502; PubMed=8354692;
RA Hess H., Heid H., Franke W.W.;
RT "Molecular characterization of mammalian cylicin, a basic protein of
RT the sperm head cytoskeleton."
RL J. Cell Biol. 122:1043-1052(1993).
CC -I- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY
CC BE INVOLVED IN SPERMATID DIFFERENTIATION.
CC -I- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE.
CC -I- TISSUE SPECIFICITY: TESTIS.
CC -I- DEVELOPMENTAL STAGE: SPECIFIC TO LATE SPERMATOGENESIS.
-----
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; Z22779; CAAR0456.1; -.
CC DR PIR; S35913; S35913.
CC DR PIR; A40713; A40713.
KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
FT DOMAIN 287 589 9 APPROXIMATE TANDEM REPEATS.
FT REPEAT 287 305 1.
FT REPEAT 306 337 2.
FT REPEAT 338 368 3.
FT REPEAT 369 405 4.
FT REPEAT 406 442 5.
FT REPEAT 443 475 6.
FT REPEAT 476 516 7.
FT REPEAT 517 547 8.
FT REPEAT 548 569 9.
FT DOMAIN 617 667 PRO-RICH.
FT SEQUENCE 667 AA; 74817 MW; CBF66EA462243D91 CRC64;

Query Match      2.8%; Score 177; DB 1; Length 667;
Best Local Similarity 20.2%; Pred. No. 0.033;
Matches 103; Conservative 89; Mismatches 180; Indels 138; Gaps 19;

QY 665 KFSHSSNPDDIATEAYEDELDMGRSGSYLNSINSAMSEHSLDPPDIR-DEUKLYAQ 724
Dl 228 KYSKSSKNNSDAVSETCKNSNVGLM-VHLGES-----DAESMEFDFWLKKNYSQ 276
QY 725 LEIYKRKMITNPNHLOKRCCK---KGLGRSIRMRITETVSRCSKEDKEDGADHT 781
Dl 277 -----NNSKPTKKDAKDAKAGKSD-----AESVDSKAKDKKNGATKDT 317
QY 782 AKGTALIRKNPPSSNGT---GKSKEETLKNRVFLSK---SHSTYDHRDQTEESSLP 835
Dl 318 KKGAKKDTSTDAESGSDSKAKGKESKKDKKDAKDAKDAASDAESGSDSKAKKSKGK 377
QY 936 TESQEE-----ETTENTSLLESLSGKLTQKLEDS-----EASTESVP 874
Dl 378 KDSKKNKKKDAKDAKDAESTDAESGSDSKAKKDKKDKKDKKDKKDAKDAKDAESTDA-- 435
QY 875 LVCKSASAHNLSEKKTGHPTSMLO-KSLSVIASAKEKTLGLAGTKTQTAGVVERTKSQ 933
Dl 436 ---ESGDSKNAKDKSKGKDKKDKKDAKDAKDAVSTDAESGDAKSK---KDSKDKK 488
QY 934 PLPKKETNRNHSNDTETKDPAPQNSNPAEPRKPKQSGIMKQQRVNPPTANSDLNPG 993
Dl 489 DLKKDQKKPKAMSKESTETESDWESKVKRDSKKDTKTA---KKATESSGAESD--- 541
QY 994 TTQMKDNFDIGECVPEVYDLTPGPVPSSEKVKOKHVSIVASEMEKNPTFLSKKSHKPK 1053
Dl 542 -----VSSKRYLKTETEFKSSDAE-----SEESLEKP- 568
QY 1054 AAEVCQSNOKRIDKAEVCLWESQGSQILEDEKL-LLSKTPVLPVPERAKEENGQ----- 1106
Dl 569 -----GSKRVDESATSTDSKKDAVEPKRGIKMPSRRTTTFKEGKKKIGTGRVPSPRE 621
QY 1107 -----PRAANVACQSEELPPK 1123
Dl 622 RPPLPCEPILPSPRVKRLCRCQMPPPPPK 651

RESULT 12
TPR_HUMAN
ID TPR_HUMAN STANDARD; PRT; 2349 AA.
AC P12270;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
-----
```



```
Db 777 IGYTCLLAACFFFAFKSRKLPENFEAKFITFSM---LIFFIVWISFIPAYASTYGV 833
Qy 634 LASRLQSDWMLLYFAHTHLTVTVTIGLLLPKFSSNN---NPRDDIATEAYEDELDM- 689
Db 834 SAVE-----VIALAASFGLLACIFFNKVYIILFKPSRNTIEVRSSTAHAFAKVAARAT 888
Qy 690 -----GRSGSYLNSINSAMSHSLDPEDIRDELKLYLAQLIYKRKKMI 734
Db 889 LRRPNITSRRKSSSLGGSTGNSPSSSTSSKSSNGEDRFPQPEROKQOQPLALTOEQOQOPL 948
Qy 735 TNNPHLOKK---RCSKK---GLG-----RSMRR-----ITEIPETV 765
Db 949 TLOPOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1008
Qy 766 SR-----QCSKDEKAGDHGTAKGTALIRKNPPSSGNTGSKKEETLKNRVFSLKKS 817
Db 1009 NRHQALLPLQCAEADSEMTIQTGLQGPVGDHPQEI-----ESPDMSPALVVSTSR 1062

RESULT 14
AC15_MOUSE
ID AC15_MOUSE STANDARD: PRT; 1131 AA.
P35601:
01-JUN-1994 (Rel. 29, Created)
01-OCT-1996 (Rel. 34, Last sequence update)
01-OCT-2000 (Rel. 40, Last annotation update)
DE ACTIVATOR 1 140 KDA SUBUNIT (REPLICATION FACTOR C LARGE SUBUNIT) (A1
DE 140 KDA SUBUNIT) (RE-C 140 KDA SUBUNIT) (ACTIVATOR 1 LARGE SUBUNIT)
DE (A1-P145) (DIFFERENTIATION SPECIFIC ELEMENT BINDING PROTEIN)
DE (ISRE-BINDING PROTEIN).
GN RFC1 OR RECC1 OR IEF-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=94089669; PubMed=8265586;
RA Burdello P.D., Utani A., Pan Z., Yamada Y.;
RT "Cloning of the large subunit of activator 1 (replication factor C)
RT reveals homology with bacterial DNA ligases";
RL Proc. Natl. Acad. Sci. U.S.A. 90:11543-11547(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94158835; PubMed=8114700;
RA Luckow B., Bunz F., Stillman B., Lichter P., Schuetz G.;
RT "Cloning, expression, and chromosomal localization of the
RT 140-kilodalton subunit of replication factor C from mice and
humans.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS;
RX MEDLINE=95386065; PubMed=7659092;
RA McGehee Habener J.F.;
RT "Differentiation-specific element binding protein (DSEB) binds to a
RT defined element in the promoter of the angiotensinogen gene required
RT for the irreversible induction of gene expression during
RT differentiation of 3T3-L1 adipoblasts to adipocytes.";
RL Mol. Endocrinol. 9:487-501(1995).
RN [4]
RP SEQUENCE FROM N.A.
RA Haque S.J.;
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-565 FROM N.A.
RC STRAIN=LAF1;
RA Lossie A.C.; Haugen B.H., Wood W.M., Camper S.A., Gordon D.F.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 354-528 FROM N.A.
```

```
RX MEDLINE=90229765; PubMed=1691767;
RA Haque S.J., Kumar A., Fischer T., Rutherford M.N., Williams B.R.;
RT "Evaluation of inter- and intramolecular primary structure homologies
RT of interferons by a Monte Carlo method.";
RL J. Interferon Res. 10:31-31(1990).
CC -!- FUNCTION: THE ELONGATION OF PRIME DNA TEMPLATES BY DNA POLYMERASE
CC DELTA AND EPSILON REQUIRES THE ACTION OF THE ACCESSORY PROTEINS
CC PCNA AND ACTIVATOR 1. THE 140 SUBUNIT BINDS TO THE PRIMER-
CC TEMPLATE JUNCTION.
CC -!- SUBUNIT: HETEROPTAMER OF SUBUNITS OF 140/145, 40, 38, 37, AND
CC 36.5 KDA THAT FORMS A COMPLEX WITH PCNA IN THE PRESENCE OF ATP.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE ACTIVATOR 1 140 KDA SUBUNIT FAMILY.
CC -!- SIMILARITY: CONTAINS 1 BRCT DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; U01222; AAA21643.1; -
CC EMBL; X72711; CAA51260.1; -
CC EMBL; U36441; AAG79698.1; -
CC EMBL; U07157; AAC52140.1; -
CC EMBL; U15037; AAB60452.1; -
CC MGI; MGI:97891; Reccl.
CC InterPro; IPR001357; -
CC Pfam; PF00533; BRCT; 1.
CC PROSITE; PS0172; BRCT; 1.
CC DNA replication; ATP-binding; Transcription regulation; DNA-binding;
KW Activator; Nuclear protein.
FT DOMAIN 399 477 BRCT.
FT NP_BIND 635 642 ATP (BY SIMILARITY).
FT ZN_FING 734 751 C2HC-TYPE (POTENTIAL).
FT DOMAIN 1104 1108 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 354 528 INTERFERON-STIMULATED-RESPONSE-ELEMENT
FT BINDING REGION.
FT CONFLICT 66 66 Y -> N (IN REF. 3).
FT CONFLICT 187 187 E -> EPDFCLSLIFGIQ (IN REF. 4).
FT CONFLICT 254 254 V -> A (IN REF. 5).
FT CONFLICT 559 559 N -> S (IN REF. 4).
FT CONFLICT 614 614 MISSING (IN REF. 3 AND 4).
FT CONFLICT 945 945 S -> N (IN REF. 1).
FT CONFLICT 1071 1071 T -> A (IN REF. 3).
FT CONFLICT 1104 1104 K -> Q (IN REF. 4).
SQ SEQUENCE 1131 AA; 125984 MW; A6F4F970A7F9EE94 CRC64;

Query Match 2.7%; Score 174.5; DB 1; Length 1131;
Best Local Similarity 23.6%; Pred. No. 0.087;
Matches 108; Conservative 76; Mismatches 157; Indels 117; Gaps 23;

Qy 761 IPETVSRQCKEDKADGHTAKGTALIRKNPPSSGNTGSKKEETLKNRVFSLKSHST 820
Db 17 VNETHV-----KNEKTKASEGTVKGGKGVKEAKVNSG-----KEDASKPKQHS-KKKRII 65
Qy 821 YDHVRDQTEESSSLPTESQBEETTE-----NSTLESLSGK--KLTQKLKEDSEAESTES 872
Db 66 YD-----SDSEETVQVKNNAKKSEKLSLSYKPGKVSQKDPVTVYVSETDE 112
Qy 873 VPLVCKSAHNLSEKKTG-----HPRTSMLOKLSLVSIAKAKETLGLACKGTOTA---- 923
Db 113 DDFVCKKRAA-----SKSKENGVTNSYLGTSNVKKNEENV-KTKNKPPLKLTPTSVDLY 167
Qy 924 -GVEERTKSKQPL--PKDKETNRNHSNSD-NTETKDPAPONSFAEPRKPKQSGIMKQ 979
Db 168 FGTESVQRSGKKMVTSKRKSSQNTDSKLNDEATAKQLQDLDEDAELERQ-----LHDE 222
Qy 980 RVNPTTANSDLNPGTOMKDNFDIGEVCPWEVYDLTPGPVPSSEKVKQHVIVASEMEKN 1039
```

Db 223 EFARTLALLDEEPKIKKARKDSEGE-----ESFSSVQDDLS--KAEKQKS 266
QY 1040 PT-----FSLKEKS-----HHKPAKE-----VCOOSNOKRI-----DRAEYCLWESQGQSI 1081
Db 267 PNKAELFSTARKTYSAPKAGHGRASEDAKQPCSAURKEACSPKASAKLALMKAKEESS 326
QY 1082 LEDEKLLISKTPTVLPRAKEENGCGOPRAANVAGQSEELPPKAVAKTENENLNQICHQE 1141
Db 327 YNETELLAAR-----RKESATEPKG-----EKTTPKTKVSPFTRKESVSPDESEK 371
QY 1142 KRTSSSEENVRGSSNNFQOP-----LTSRAEVC 1172
Db 372 KRTNYQ-----AYSYLNREGPKALGSKEIPKGAENC 403

RESULT 15

NFH_HUMAN STANDARD; PRT; 1020 AA.
P12036;
01-OCT-1989 (Rel. 12, Created)
01-OCT-1989 (Rel. 12, Last sequence update)
01-OCT-2000 (Rel. 40, Last annotation update)
DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
GN NEFH OR NFH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88328981; PubMed=3138108;
RA Lees J.F., Sheidman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;
RT "The structure and organization of the human heavy neurofilament
subunit (NF-H) and the gene encoding it.";
RL EMBO J. 7:1947-1955(1988).
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLET K-S-P, NFH IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15306; CAA33366.1; -
DR EMBL; X15307; CAA33366.1; JOINED.
DR EMBL; X15308; CAA33366.1; JOINED.
DR EMBL; X15309; CAA33366.1; JOINED.
DR PIR; S00979; QFHUH.
DR MIM; 162230; -
DR InterPro; IPR001664; -
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Heptad repeat pattern; Coiled coil; Neurone;
KW Phosphorylation.
FT DOMAIN 1 100 HEAD.
FT DOMAIN 101 413 ROD.
FT DOMAIN 414 1020 TAIL.

FT DOMAIN 101 132 COIL 1A.
ET DOMAIN 133 145 LINKER 1.
FT DOMAIN 146 244 COIL 1B.
FT DOMAIN 245 266 LINKER 12.
FT DOMAIN 267 288 COIL 2A.
FT DOMAIN 289 292 LINKER 2.
FT DOMAIN 293 413 COIL 2B.
SQ SEQUENCE 1020 AA; 111780 MW; 1177C9DCB3DCF1D4 CRC64;
Query: Match 2.7%; Score 174; DB 1; Length 1020;
Best Local Similarity 20.6%; Pred. No. 0.081;
Matches 101; Conservative 79; Mismatches 229; Indels 82; Gaps 17;
QY 754 IMRRITEIPETVSRQSKEDKAGDHGTAKGTALIRKNPP-ESSNGTGSKEETLKNRVF 812
Db 466 VTEEVTEEEKEAKEEGEGEGEEAEAGGEEETKSPPAEEAAASPEKEAKSPVKEEK 525
QY 813 SLAKSHS-----TYDHRDQTESSSLTESOEETTE-NTLSLSGKKLT 858
Db 526 SPAEAKSPEKEEAKSPAENVKSPKAKSPAEEAKSPKAEKSPKAEKSPAENVKSPKAK 585
QY 859 QKLKEDSEAESESTESVPLVCKSAHANLSSEKKTGHPRTSMLOKSLSVIA--SAKETLGL 916
Db 586 SPAKEEAKSPAEEAKSPAENVKSPKAEKSPAEEAKSPAENVKSPKAEKSPAENVKSPK 645
QY 917 AGKTQTAGVEER-----TKSQKPLPKDRETNRNHNSNDNTETKDPAPQNSNPAEPRKP 970
Db 546 AKSPEKAKSPEKEEAKSPAENVKSPKAEKSPAENVKSPKAEKSPAENVKSPKAEKSP 705
QY 971 QKS-GIMKQORVNPNTANSDLNPGTTQMDNFDIGEVCPEWVYDLTPGPVPSKVKOKHV 1029
Db 706 EKAKSPVKEEAKSPAENVKSPKAEKSPAENVKSPKAEKSPAENVKSPKAEKSPAENV 736
QY 1030 SIVASEMEKNPTFSLAEKSHHKPKAAEVCOOSNQKRIDKAEVCLWESOGOSILEDEKLLI 1089
Db 737 KSPVKEEAKSPAENVKSPKAEKSPAENVKSPKAEKSPAENVKSPKAEKSPAENVKSPK 784
QY 1090 SKTPVLPRAKEENGCGOPRAANVAGQSEELPPKAVASK--TENENLNQIGHOEKKTSS 1147
Db 785 AKSPVKEEAKSPAENVKSPKAEKSPAENVKSPKAEKSPAENVKSPKAEKSPAENVKSPK 843
QY 1148 EENVRGSSNNFQOPLTSRAEVCPEWETP---AQPNAGRSVALPASSALSANKIAGP 1204
Db 844 EEKAPATPKT---EEBKDSKKEEAP-KKEAPKPVKEEKKEPAVEKPKESKVEA----- 892
QY 1205 RKEEINDSEKV 1215
Db 893 KKEEEDKXKV 903

Search completed: November 1, 2001, 10:20:45
Job time: 378 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 10:13:32 ; Search time 61.87 Seconds
(without alignments)
2598.200 Million cell updates/sec

Title: US-09-775-181-2
Perfect score: 6382
Sequence: 1 MGAWAYPELLICLLLAQLGLG.....LSANKIAGPRKEIWDSEFKV 1215

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Archived: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_16.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_unclassified.*
- 13: sp_vertebrate.*
- 14: sp_virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2811	44.0	544	Q9ULT3	Q9ult3 homo sapien
2	1532.5	24.0	317	Q9GKT3	Q9gkt3 macaca fasc
3	571.5	9.0	669	Q9VR40	Q9vra40 drosophila
4	307	4.8	1677	Q9VKA3	Q9vka3 drosophila
5	281	4.4	264	Q9WLP3	Q9wlp3 drosophila
6	279.5	4.4	433	Q45500	O45500 caenorhabdi
7	254	4.0	1221	Q9Y133	Q9y133 drosophila
8	238.5	3.7	176	Q9WLP4	Q9wlp4 drosophila
9	227.5	3.6	1305	Q9WPS7	Q9wps7 drosophila
10	225	3.5	5327	Q96891	O76891 drosophila
11	210.5	3.3	1199	Q9EPV6	Q9epv6 mus musculu
12	209	3.3	1160	Q9ESC8	Q9esc8 mus musculu
13	205.5	3.2	3111	Q9VH10	Q9vhl0 drosophila
14	205	3.2	1183	Q9W3D1	Q9w3d1 drosophila
15	204.5	3.2	6632	Q17362	Q17362 caenorhabdi
16	204.5	3.2	6642	Q01761	O01761 caenorhabdi
17	201.5	3.2	700	P82179	P82179 canis famil
18	201.5	3.2	1200	Q25388	Q25388 loligo peal
19	200	3.1	2081	Q9LH98	Q9lh98 arabidopsis

20	198.5	3.1	1812	5	Q9VZD9	Q9vzd9 drosophila
21	198	3.1	877	4	Q9UGT0	Q9ugt0 homo sapien
22	198	3.1	6677	5	Q9N435	Q9n435 caenorhabdi
23	195	3.1	1338	11	Q9JIX8	Q9jix8 mus musculu
24	194.5	3.0	3488	5	P91257	P91257 caenorhabdi
25	193	3.0	1481	5	Q9VFR6	Q9vfr6 drosophila
26	190	3.0	1390	5	O77033	O77033 dictyosteli
27	189	3.0	808	4	Q9UK88	Q9uk88 homo sapien
28	187	2.9	1046	5	P91176	P91176 caenorhabdi
29	186	2.9	522	10	Q9FXB5	Q9fxb5 arabidopsis
30	185	2.9	3201	5	Q9W0U2	Q9wou2 drosophila
31	183	2.9	852	10	Q9SZ55	Q9sz55 arabidopsis
32	183	2.9	976	5	Q9V485	Q9v485 drosophila
33	182.5	2.9	1163	4	Q9UHB7	Q9uhb7 homo sapien
34	181.5	2.8	1698	2	Q9LC00	Q9lc00 staphylococ
35	181	2.8	2151	5	Q9V1L3	Q9v1l3 drosophila
36	181	2.8	2151	5	Q9VPL6	Q9vpl6 drosophila
37	181	2.8	2151	5	Q9NG79	Q9ng79 trichomonas
38	181	2.8	5322	5	Q9NI64	Q9ni64 drosophila
39	179.5	2.8	423	5	Q9VVH1	Q9vvh1 drosophila
40	179.5	2.8	1230	5	Q20626	Q20626 caenorhabdi
41	179.5	2.8	3261	4	Q9Y556	Q9y556 homo sapien
42	179	2.8	2031	13	Q9PVV4	Q9pvv4 xenopus lae
43	178.5	2.8	1020	5	Q27104	Q27104 trichomonas
44	178.5	2.8	1151	5	Q9VAY4	Q9vay4 drosophila
45	178	2.8	2451	5	Q9VG05	Q9vg05 drosophila

ALIGNMENTS

RESULT 1

ID Q9ULT3 PRELIMINARY; PRT; 544 AA.
AC Q9ULT3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE KIAA1136 PROTEIN (FRAGMENT).
GN KIAA1136.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC [ISSUE=BRAIN;
RX MEDLINE=20039618; PubMed=10574461;
RA Hirose M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;
RT "Characterization of cDNA clones selected by the GeneMark analysis
from size-fractionated cDNA libraries from human brain.";
RL DNA Res. 6:329-336(1999).
DR EMBL; AB032962; BAA86450.1; -.
FT NON_TER 1
SQ SEQUENCE 544 AA; 60195 MW; DA1175E7926041F7 CRC64;

Query Match 44.0%; Score 2811; DB 4; Length 544;
Best Local Similarity 99.8%; Pred. No. 8.4e-196;
Matches 543; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 672 NNPDDIATEAYEDELDMGRSGSYLNSSINSANSEHSLDPEDIRDELKLYAQLEYKRR 731

DB 1 NNPDDIATEAYEDELDMGRSGSYLNSSINSANSEHSLDPEDIRDELKLYAQLEYKRR 60

QY 732 KMTNPNHLOKKCKSKGLGRSITMRITETPEVSRQCSKEDKGDHGTAKTALIRKN 791

DB 1 KMTNPNHLOKKCKSKGLGRSITMRITETPEVSRQCSKEDKGDHGTAKTALIRKN 120

QY 792 PPSSNGTGKSKETLKNRVFSLKSHSTYDHRDPTERSSSLPTESQRETTENSTLES 851

DB 121 PPSSNGTGKSKETLKNRVFSLKSHSTYDHRDPTESSSSLPTESQRETTENSTLES 180

```
QY 852 LSGKKLTOKLKEDSEAEESTESVPLVCKSASAHNLSSEKKTGHPRPTSMLOKLSVSIASAKE 911
|||||
Db 181 LSGKKLTOKLKEDSEAEESTESVPLVCKSASAHNLSSEKKTGHPRPTSMLOKLSVSIASAKE 240
|||||
QY 912 KTLGLAGKTQTAGVEERTKSKPLPKDKETNRHNSNSDNTETKDPAPONSNAEPRKPQ 971
|||||
Db 241 KTLGLAGKTQTAGVEERTKSKPLPKDKETNRHNSNSDNTETKDPAPONSNAEPRKPQ 300
|||||
QY 972 KSGIMKQORVNPPTTANSNDLNFQTKMDNFDIGVEYCPWEVYDLTPGVPSESKVKHYSI 1031
|||||
Db 301 KSGIMKQORVNPPTTANSNDLNFQTKMDNFDIGVEYCPWEVYDLTPGVPSESKVKHYSI 360
|||||
QY 1032 VASEMEKNPTSLSEKSHHKPKAAEVCQSNQKRIKAEVCLWESQSGSILEDEKLLISK 1091
|||||
Db 361 VASEMEKNPTSLSEKSHHKPKAAEVCQSNQKRIKAEVCLWESQSGSILEDEKLLISK 420
|||||
QY 1092 TPVLPERAKEENGOPRAANVCAGQSEELPPKAVASKTENENLNOIGHOEKKTSSSEENV 1151
|||||
Db 421 TPVLPERAKEENGOPRAANVCAGQSEELPPKAVASKTENENLNOIGHOEKKTSSSEENV 480
|||||
QY 1152 RGSYNSSNNFQOPLTSRAEVCWPEFETPAQPNAGRSVALPASSALSANKIAGPRKEETWD 1211
|||||
Db 481 RGSYNSSNNFQOPLTSRAEVCWPEFETPAQPNAGRSVALPASSALSANKIAGPRKEEWD 540
1212 SFKV 1215
|||||
Db 541 SFKV 544
|||||

RESULT 2
Q9GKT3 ID Q9GKT3 PRELIMINARY; PRT; 317 AA.
AC Q9GKT3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 34.7 KDA PROTEIN.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CEREELLUM CORTEX;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AB052146; BAB19001.1; -
Hypothetical protein.
SEQUENCE 317 AA; 34746 MW; 26C0DB80505EFC4 CRC64;

Query Match 24.0%; Score 1532.5; DB 6; Length 317;
Best Local Similarity 93.4%; Pred. No. 2.7e-103;
Matches 297; Conservative 3; Mismatches 17; Indels 1; Gaps 1;

QY 898 MLOKLSVSIASAKEKTGLAGKTQTAGVEERTKSKPLPKDKETNRHNSDNTETKDPQ 957
|||||
Db 1 MLOKLSVSIASAKEKTGLAGKTQTAGVEERTKSKPLPKDKETNRHNSDNTETKDPQ 60
|||||
QY 958 PQNSNPAEPRKPQKSGIMKQORVNPPTTANSNDLNFQTKMDNFDIGVEYCPWEVYDLTPG 1017
|||||
Db 61 PQNSNPAEPRKPQKSGIMKQORVNPPTTANSNDLNFQTKMDNFDIGVEYCPWEVYDLTPG 120
|||||
QY 1018 PVPSESKVKHYSIVASEMEKNPTSLSEKSHHKPKAAEVCQSNQKRIKAEVCLWESQ 1077
|||||
Db 121 PVPSESKVKHYSIVASEMEKNPTSLSEKSHHKPKAAEVCQSNQKRIKAEVCLWESQ 180
|||||
QY 1078 GOSTLEDEKULLISKTPVLPERAKEENGOPRAANVCAGQSEELPPKAVASKTENENLNOI 1137
|||||
```

```
Db 181 QOSTLEDEKULLISKTPVLPERAKEENGOPRAANVCAGQSEELPPKAVASKTENENLNOI 240
|||||
QY 1138 GHQEKKTSSSEENVGYSYNNFQOPLTSRAEVCWPEFETPAQPNAGRSVALPASSALS 1197
|||||
Db 241 GHQEKKTSSSEENVGYSYNNFQOPLTSRAEVCWPEFETPAQPNAGRSVALPASSALS 299
|||||
QY 1198 ANKIAGPRKEEWDSEFKV 1215
|||||
Db 300 ANKIAGPRKEEWDSEFKV 317
|||||

RESULT 3
Q9VR40 ID Q9VR40 PRELIMINARY; PRT; 669 AA.
AC Q9VR40;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CG11923 PROTEIN.
GN CG11923.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flooker A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AF003575; AAF50967.1; -
DR FlyBase; FBgn0031642; CG11923.
DR InterPro; IPR000337; -
DR Pfam; PF00003; 7tmL3.1.
SQ SEQUENCE 669 AA; 75242 MW; FFE5721445DAB5F6 CRC64;
```


Db 1360 FKQGDQKWDQ-----KAKVRSITASTSFLNGVLVPESPDPOLYQENELKEQVOKLAHQ 1412
QY 725 LEIYRKMKMTNNPHLQKRCCKGLGRSIRMRITEIPETVSRQCSKEDKEGADHGHTAK 784
Db 1413 IEFMKTVHQMNRHLK-----PKPGGYFITSTSFQAPYSKNTVSTAQTQTKDENSVDK 1468
QY 785 TALIRKNPESGNTGSK--BETILKNRVFSKLKSHSYDHRDQTERSSSLPTSQEBEE 842
Db 1469 CSI-----EIDGGGRTLVLEDATE---FNF-----HJAD--TEKGLIVQEDERP 1509
QY 843 TTENSTLESLSCKKLTQKLDKEDSEASTESVPLVCKKSASAHNLSEKKTGHPTSMLOKS 902
Db 1510 SDEEAALIAQRRFLAPILDS-----LNLYYQLNDLDDTEHVRHIQHTVAQ 1556
QY 903 LSVIASAKETLGLACKTQTACVBERTKSQPLPKDKETNRNHSNDTETKDPAPQNSN 962
Db 1557 MNDLTSSEETM-----VTVQNSPSPPPVGVVELLPISSDSST-----ASSSLY 1600
QY 963 PAEPRKPKQSKS--MKQVRNPPTTANSLDNLPGTQMK 998
Db 1601 AHTPSAPHPSCVLLMQLNPLESPLLSGSDAITITQVE 1638
RESULT 5
ID Q9WIP3 PRELIMINARY: PRT: 264 AA.
AC Q9WIP3
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CG18678 PROTEIN.
GN CG18678.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Branton R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flook C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M.C., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.-N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
DR EMBL; AE003461; AAF47013.1; -.
DR FLYbase; FBgn0040664; CG18678.
DR InterPro; IPR000337; -.
DR Pfam; PF00003; 7tm3.1.
DR PRINTS; PR00248; GPCRMR.
SQ SEQUENCE 264 AA; 29532 MW; CAC5623C52703642 CRC64;
Query Match 4.4%; Score 281; DB 5; Length 264;
Best Local Similarity 28.5%; Pred. No. 15e-12;
Matches 65; Conservative 49; Mismatches 96; Indels 18; Gaps 5;
QY 472 LYFEPSTFCILLRWARLLGFATVYGVTVLKLHRLVKVFLSSTAQRIPYMTGGRVMRLA 531
Db 1 MYPNPNLYTCTARIWLREIGFSLTYGALMLKTWRTSVIFRVSRAVK-ITDAALLKRLG 59
QY 532 VILLVFWFLIGWTSVCONLEKQISL--IGCKTSDHLIFNMCLIDRWDMYTAVEFLF 589
Db 60 ICAGI-----GTCLLVRTLVPDPVVGVTADDLKAFKLCTDMWDYFTTSNEVL 110
QY 590 LMGVYLCYAVRTVPSAFEPYMAVAVHNELIISAFTIRFVLASRLQSDMLMLYFA 649
Db 111 LAWGVRLCIMVKKAPSEFSESRFTSMAYINEFLTCLFNLVSMFLQSPANPDLLIIFFC 170
QY 650 HTHLTVTVTIGLLL---IPKFSHSSNNRPDDIATAYEDELDMGRSGS 694
Db 171 HTQLVTVELLALIFGSKIPTLTQVVIVLVR---SGKSHQENIGMGPKAS 215
RESULT 6
ID O45500 PRELIMINARY: PRT: 433 AA.
AC O45500;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)
DE F39B2.8 PROTEIN.
GN F39B2.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Dobson R.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans";
RL Nature 368:32-38(1994).
DR EMBL; Z92834; CAB07388.1; -.
SQ SEQUENCE 433 AA; 49673 MW; B35F0A9C23DA096 CRC64;
Query Match 4.4%; Score 279.5; DB 5; Length 433;


```
Db 628 AYKGLLVFAGLANETRVHSIPALNDSKHIGFVSVINVFITCLAGAAISLVLSDDRDLVF 687
Qy 643 MLMLYPATHLHVTVITIGLLIPKFSHSSNNPRDDIATEAYEDELDMRGSGYLNSSINS 702
Db 688 VLLSFF--IIFCTATLCLVFPKVLKRNQGVVKKVRVTRLPMSKNGR----- 737
Qy 703 ANSEHSLDPEDIRDELKLYLAOLEYKRYKKMTIN-NPHLQKRCCKKGLGRSIRMRITEI 761
Db 738 -----RDS-----SVCELEQRURDKVNTNCRPRKALMEKENELQALIRKLG-- 778
Qy 762 PETVSRQ-----CSKEDKEGADHGTAKTALIRKNPPESSNTGKSEETLKNRVFSLK 815
Db 779 PE--ARKWIDGVTCTGGSNVSGSELEPILNDDIVRLSAPPVRREMPSTTEVTWTSVDSVT 836
Qy 816 KSHSYDHY--VRDOTEESLSLPTESOBETTEN-----STLESLSGKKLTKQLKEDSE 866
Db 837 STHVEMDNSFVSQSTVMAPLSPKPKKOSIVEHSHAPPTMMQPIQOOLQHLQHQHQ 896
Qy 867 AESTESVPLVCKSAHNLSSSEKKTGHPTSMLOKSLVSIAIAKKTGLAGKTOTAG-- 924
Db 897 MQQHQ-----LQOQHQHQOQOQOQOHHRLHLEKRNVSVAQTDD---NIGSITSTACKR 948
Qy 925 -----VZEETK-----SQPLPKDKETNRHNSNSNTETKDPAPONSNAPEEP 967
Db 949 SGGDCSSMRERQSTASRHYDSGQSTPTARPKYSSSHRNSNSTNISTOSELSNMCPSHPK 1008
Qy 968 RKPQSGIMKQQRVNTPTANSILNCPCTOMKDNFDIGECWPEVVDL-----TPGPVP 1020
Db 1009 STP-----AVIK-----TPTASDHRTSGNSALKNFVVSQSLDWTHTLSIAKORQSPRYA 1061
Qy 1021 SESKVQKHSIVASSEMKNPT-----FSLKEKSHHKPKAAEVCOOSNOKRIDKAEVCLW 1074
Db 1062 SPQRCAEHGHGCMYDNPNTTPIQRSVSEKRNKRHRPKP-----QKGTVC-- 1107
Qy 1075 ESOGOSILEDEKLLSKTPVLPERAKEENGOPRAANCAGOSELPKAVASKTENENL 1134
Db 1108 QSETOSERDRPPNSQPCVQPRKVS-----RSSNIQHAHHSSPNVAPDK---QRS 1157
Qy 1135 NQIGHQEKK---TSSEENVRGYSNNSNFFOPLTSRA-----EVPK 1173
Db 1158 RQRGQODSIYCASSETELLEGETAILPIFRKLLTEKSPNYGRSVAQGSCP 1209

RESULT 8
Q9W1P4
ID Q9W1P4 PRELIMINARY; PRT; 176 AA.
AC Q9W1P4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
CG18679 PROTEIN.
CG18679.
Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottlie P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
```

```
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount D.R., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003461; AAF47012.1;
DR FlyBase: FBgn0040663; CG18679.
DR InterPro: IPR000561;
DR SMART: SM00001; EGF-like: 1.
SQ SEQUENCE 176 AA; 19661 MW; 524E4E2C48CB7D33 CRC64;

Query Match 3.7%; Score 238.5; DB 5; Length 176;
Best Local Similarity 31.5%; Pred. No. le-09;
Matches 52; Conservative 25; Mismatches 55; Indels 33; Gaps 4;

Qy 301 RGVKVDINLQKVIDQCSSD-----GWFSCTHKCHLNNSCMPKIGLGFVLGAYE 351
Db 9 RGTSGIDIDLRVDDIDQCPQRHTPCTKRPLNIFACTDKCKORTMCEAIMGLGFRGSYK 68
Qy 352 CICKAGFYHGVLPVNNFRRGPDQIHSGSTKV-----SEAYVCLPCRE 397
Db 69 CICKRGFYFPDIVSQHKF-----FNGSLLEEYKMLGKNSTYNSSEYECLPCAE 120
Qy 398 GCPCADDSPCFVQEDKYLRLLAIISFQGLCMLLDFVSLVYVHF 442
Db 121 GDCSCDSSPCIALNWPMTSILAL--ACIVIGLLPPAAWFTFR 163

RESULT 9
Q9VPS7
ID Q9VPS7 PRELIMINARY; PRT; 1305 AA.
AC Q9VPS7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CG3022 PROTEIN.
GN CG3022.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottlie P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
```

Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
April J.F., Agayani A., An H.-J., Andrews-Prannkocch C., Baldwin D.,
Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Fosler B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003588; AAF51465.2; -
DR FlyBase; FBgn0031275; CG3022.
DR InterPro; IPR000005; -
DR InterPro; IPR000337; -
DR InterPro; IPR001828; -
DR Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 2.
DR PRINTS; PR00248; GPCRMR.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; UNKNOWN_1.
SQ SEQUENCE 1305 AA; 143588 MW; A9C3C85307650450 CRC64;

Query Match 3.6%; Score 227.5; DB 5; Length 1305;
Best Local Similarity 18.8%; Pred. No. 1.2e-07;
Matches 209; Conservative 171; Mismatches 381; Indels 351; Gaps 52;

QY 181 DLSLAP-APQVFLQATRESRILQDLSNAPH-----LANATLETFWHLRRKWRP 232
DB ESMGAPWPPQRTACSNHLEQLAVENLIVSTUNSVGNVSYSLNHNMFNSQLRKOSA 442
QY 233 HLHRRGNPGRLGHSW-----RRKDLGDKSHFKWSPPLYECNGSY 277
DB 443 QFH-----GODFGSGYGRISATATQSDRRRRRGVGTSGHLPFAISQYAPQTY 496
QY 278 KPGMLVTLSSAIYGLQNLVPEFRGVKMDINLQKVIDQCSDDGWFSGTHKCHLNSEC 337
DB 497 DAVNAIAL-----RAAEHW-----RRNEEQ 519
QY 338 MPIKGLGFVLG--AYECIKAGFVH-PCVLPVNNFRRRGPDQIHSGST-----KDYSEE 388
DB 520 SKLGDFTYRSDMAWEFLQMGKHLFCVSGPVSF--SGPDR--VGTTAFVQIQRGLEP 575
QY 389 AYVCLPCREG---CPFC-----ADDSPCFVQEDKYLRLAIIISF 423
DB 576 VALYYPATDALDFRCPRCPKRWKHSQGVPIAKRVFKLRVATIAPL-----AFYTIATLSS 630
QY 424 QGLCMLLDFVSMVYVHFRKAKSIRASGLJLITLFGSLLLYFPVVLVYFEPST-----478
DB 531 VGIALAIAFAFNL--HFRKLKAIKLSPKLSNTAVGCIFFVATVILGLDHSITLPSAE 688

QY 479 --PRCILLRWARLL--GFATVYGVTVTLKLRVLKLVFLSRTAQ-----RIPYMTGG 524
DB 689 DSFATVCTARVYLLSAGSLAFSGMFAKTYRVHRIF--TRTGSVFKDKMLQDIOILLVGG 747
QY 525 RVMRMLAVILLVWFVFLIGWTSVVCQNLKQISLIGOGKTSDDLIF-----NMCL---IDR 577
DB 748 ---LLIVDALLVLVWVTDPMERHLNLTLEIS-----ATDRSVVYQPOVEVCRSQHTOT 799
QY 578 WDMYTAVEEFLLMGYLYCYAVRTVP-SAFHEPRYMAVAVHNELIISAIFHTIRFVLAS 636
DB 800 WLSVLYAYKGLLVGVYMAWETRHVKIPALNDSQYIGSVSVVITSAT-----VVVLA 854
QY 637 RLASDWMMLYFANTHL---TVVTIGLLIPKFSHSSNNPRDDIATEAYEDELDMGRSG 693
DB 855 NLISERVTLAFITITLITLSTTTATLCLLFPK-----887
QY 694 SYLNSSINSWSHS--LDP-----EDIRDELKLYAOLEIVKXKMI 734
DB 888 -----LHDJWARDIIPVHSMGLAMECNTRFRVVDREIQYRVEQNRVYKKE---938
QY 735 TNNPHLQKKRCKGLGRSIRMRITETPVSQCKSEKDEGADH-----GTAKGT 785
DB 939 -----IQALDAEIRKLERLLESGLITTTTSSSTSL--LTGGHKLKPELTVTSGISQTP 991
QY 786 ALIRKNPSSSGNTGKSEETLANRVFSL-----KXSHSTYDHRVDQTESSSLPTESQ 839
DB 992 AASKNRTPSISG-----ILPNLLSVLPVPIPRASWPSAEYMQIPMRRSVTFASQPQ 1043
QY 840 EETTENSTLESLSGKKLTOKLKDSEASTESVPLVCKSAS-----AHNLSSEKKTG 892
DB 1044 LEE-----ACLPADQLINLRLAHQOATEAKTG 1070
QY 893 -----HPRTSMLOKSLSVASAKEKTLGLAGKTOTAGVEERTKSQKPLPKDKETNRN 944
DB 1071 LINLRGIFSRSTSSNKG-STASLADQKGLKAPKSH-MGLFTEL-----IPSSQATSCN 1123
QY 945 HSNSDNTETKDPQNSNPABEPRKPKOKSGIMKOORVNPPTANSDLNPCTOMKD-NF-D 1002
DB 1124 AIYNNPNQDSIPSEASHPNHLPKPIHRGSL-----TKSGTHLDHLT---KDPNPLP 1173
QY 1003 IGEVCPMEVYDLPVGPVPSKVKQKHSIVASEM-----EKNPTF-----SLKEKSHHK 1051
DB 1174 IPTISGGEQDQTLG-----GKYVKLLETVNFOLPSNRRPSVVQPPSRVRVRS 1225
QY 1052 PK-----AAEVCCQSNQKIDKAEVCLWESQSGSILDEKLLISKTPVLPRAKEE--NG 1104
DB 1226 PRPHRLILPTCSLS-----ALAESEDRP--CDSTSLSCSKSIPRISLQOATSG 1273
QY 1105 GOPRA-----ANVCAGQSELPKPAVASKTE 1130
DB 1274 GTWKSMTAGKSRSLGDSQEEQAPANGTE 1305
RESULT: 10
076891
ID 076891 PRELIMINARY; PRT; 5327 AA.
AC 076891;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)
DE EG:49E4.1 PROTEIN.
GN FUTS4 OR EG:49E4.1 OR CG3064.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Insecta; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Phylorhoda; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN NCBI_TaxID=7227;
RP SEQUENCE FROM N.A.
RA Papagianakis G., Spanos L., Siden-Kiamos I., Louis C.;
RT "Sequencing the distal x chromosome of Drosophila melanogaster.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

```

RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031128; CAA20006.1;
DR FlyBase; FBgn0025392; futsch.
SQ SEQUENCE 5327 AA; 575942 MW; FEFPE23A118FF38A CRC64;

Query Match 3.5%; Score 225; DB 5; Length 5327;
Best Local Similarity 21.5%; Pred. No. 1.3e-06;
Matches 128; Conservative 91; Mismatches 277; Indels 100; Gaps 22;

QY 668 SHSSNNPRDDIATEAYEDELDMGRSGSYLNSSINSAWSEHSLDP----EDIRDELKLYA 723
DB 1716 AESAKDGADDLKELSRPESTTQSKAGSIKDEKSPLASSEASRPASVAESVKADEAKSKE 1775

QY 724 QLEIYKRWMTNNPHLQKRCSSKGLGRSINRRITEIPETVSROCKEDKAGDHGAK 783
DB 1776 E-----SRRESVAESKPLPSKEAS-----RPASVAESIKDEAKSKEESRRESVAE 1821

QY 784 GTALIRK--NPPE-----SGNTGKSKEETLKNRVF-----SLKSHSTYDHRDQT 828
DB 1822 KSPPLSKASRPASVAESIKDEAKSKEESRRESVAESKPLPSKEASRPASVAESIKDEA 1881

QY 829 EESSLPTESQEEETENTLES-----SGKKLTOKLKEDSEAES--TESVPLVC 877
DB 1882 EKSKE---ESRRESVAESKPLPSKEASRPASVAESIKDEAKSKEESRRESVAESKPLPS 1938

QY 878 KSA-----AHNLSE---KKTGCHPTSMLOKS-----LSVIAKAKETLGLAGKTOTA 923
DB 1939 KASRPASVAESIKDEAKSKEESRRESVAESKPLPSKEASRPASVAESIKDEAKSKEE 1998

QY 924 GVEERTKSKPLPKOKETNRNNSNDTETKDPAPONSNPABPRKPKOKSGIMKQORVNP 983
DB 1999 SRRESVAESKPLP--SKEASRPASVAES--IKDEAKSKEESRRESVAESKPLPSKEASRP 2055

QY 984 TTANDLNPQTOMKDNFDIGECVPEVYDLPGVPVPSKVKOKHVSIVASMEKNPTFS 1043
DB 2056 ASVAESIKDEAKSKE-----ESRRESAAKSPPLPSKEASRPASVAESVKADEAKSKEES 2110

QY 1044 LAKSHHKPKAEV--COOSNOKRIDKAEVCLWESOGOSILEDEKLLIKTPVLPRAKKEE 1102
DB 2111 RESMAESKASGSIKDGOSPLKESVRP-----ESVAESKVD-----PVKSKEP 2154

QY 1103 NGQOPRAANVCA--GQSEELP-----PKAVASKTENENLQIGHOEKTKTSSEENVR 1152
DB 2155 SRRESVAGSVTADSDQDQSPLESKASRPESVVDVSKDEAKQESRRESKTESVIPPKA 2214

QY 1153 GSYNSNNFOOPLTSAEVCVPEFETPAQPNAGRVALPASSALSANKIAGPRKEE 1208
DB 2215 KDDKSPKEVLPV--SWTETIREDADQPMKPSQAE-----RRESIAESIKASSPRDEK 2266

RESULT 11
Q9EPV6 PRELIMINARY; PRT: 1199 AA.
AC Q9EPV6;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR TYPE 1.
GN GLUR1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=BRAIN;
RX MEDLINE=20050857; PubMed=10581402;
RA Zhu H., Ryan K., Chen S.;
RT "Cloning of novel splice variants of mouse mGluR1.";
```

```

RL Brain Res. Mol. Brain Res. 73:93-103(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=BRAIN;
RA Zhu H., Ryan K., Chen S.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF320126; AAG41991.1;
KW Receptor.
SQ SEQUENCE 1199 AA; 133386 MW; C13F2E9CA83EA024 CRC64;

Query Match 3.3%; Score 210.5; DB 11; Length 1199;
Best Local Similarity 18.8%; Pred. No. 1.8e-06;
Matches 165; Conservative 119; Mismatches 330; Indels 263; Gaps 31;

QY 357 GFYHPGVLPVNNFRRRGPDHSGSTKDYSE-----AYVCLPCREG- 398
DB 498 GTWHEGVNIDDIKIQ---MNKSGMVRVSCPECLKQGIKVRKGEVSCCWTACKENE 554

QY 399 -----CPFC-----ADDSPCFVQEDKYLR-----LAISFQGLCMLDVFVSLV 438
DB 555 FVQDEFTCRACDLGWPNAELTGCEPIIRYLEWSDIESIINAFSCLGILVLFVTLIF 614

QY 439 YHFRKASIRASGLILLETILPGSLLLPVVPVILYFEPSTFCILLRWARLLGPAFVYGT 498
DB 615 VLYRDPVVKSSRELXYILLAGIFLGVCPPTLIATPTTSCYLQRLVGLSSAMCYSA 674

QY 499 VTLKLRHLV-----FLSRTAQRIPYMTGGVRVRLAVILLVWFVFLIGW 544
DB 575 LVTKNRIARILAGSKKIKCRKPRFSAWAQVITIASILSVOLTLVTLII----- 726

QY 545 TSSVCNLEKQISLIGOGKTSDLIFNMLCLDRWDYMTAVA-EFLFLMGVYLCYAVRTV 603
DB 727 -----NEPMPILSYPSIKE--VYLICNTSLNGVAPGVYNGLLIMSCITYAEKTRNV 777

QY 604 PSAPHEPRYMAVAVINELIISAIFHTIRFVLASRLQSDWMLMLYFAHHTLVTVTIGLL 663
DB 778 PANFENEAKYIAFTVYTTCIIWLAFAVPIYF-----GSNYKIITTCFAVSLSVTALGCMF 831

QY 664 IPK-----FSHSSNNPRDDIATEAYEDELDMG-----RSGSYLNSSINSINSAWSEHSLDP 711
DB 832 TPKWIIIIIAKERNVRSAPT---SDVVRMHVGDGKLPKRSNTFLN----- 874

QY 712 EDIRDELKLYAQLIYKRRKMITNPHLQKRC-----KKGLGRSIRRIITEIPET 764
DB 875 -----IFRRKPGAGNANSNGSVSWSEPGGQAQPKQHVQWQRLSVHVKT 919

QY 765 VSRQSKEDKCADHGTAKTALIRKNPESSGNWTKSKEETLKNRVFSLKSHSTYDHY 824
DB 920 NETACNQ-----TAVIK---PLTKSYOGSGKSULT-----FSDASTKLYN-- 956

QY 825 RDQTESSSLPTESQEEETENTLESLSGKLTQKLEDSEAESESVPLVCKKSASAHN 884
DB 957 ---VERGOYFNSLQ-----PSQOPSWVHR 980

QY 885 LSSEKKTGHPRTSMLOKSLSVIASAKETLGLAGKTOTAGVEERTKSKQPLPKDKETNRN 944
DB 981 RGPVAVATTPPLPHL-----SAEETPLFADSVIPKGL-----PPPLPQ----- 1019

QY 945 HNSDNTETKDPAPONSNPABPRKPKOKSGIMKQORVNPVTANSDL-----NPGTQ 996
DB 1020 -----QQQQPPPPPPPPQPPQKSLMDLQGVVTFNFGSGIPIDFHVLAVLAGPGT-- 1063

QY 997 MKDNFDIGEVCPWE-VYDLTPGPVPSKVKOKHVSIVASEMEKNPTFSLEKSHHKPKAA 1055
DB 1064 -----PGNCLRSLYPPPPPPQ-----HLQMLPLQLS-----TFREPISPP--G 1100

QY 1056 EYCOOSNOKRIDKAEVCLWESQ-----QSTLEDEKLLISKTVPVLPRAKEENGQOPRAANV 1112
DB 1101 EDDDDSSEREKLLQEFVYREGNTEEDLEBEDLPAASKLTPEDSPALTPSPSPRDSV 1160

QY 1113 CAGQSEELPP--KAVASKTENENLNOIHOEKKTSSS 1147
DB 1113 CAGQSEELPP--KAVASKTENENLNOIHOEKKTSSS 1147
```


RESULT 15

Q17362 Q17362 PRELIMINARY; PRT; 6632 AA.
AC Q17362;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE UNC-89 PROTEIN.
GN UNC-89.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
MDLINE=96180278; PubMed=8603916;
RT Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT assembly, encodes a giant modular protein composed of Ig and signal
RT transduction domains."
RL J. Cell Biol. 132:835-848(1996).
CC -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR EMBL; U33058; AAB00542.1; -.
DR HSSP; P56276; ITLK.
DR InterPro; IPR000219; -.
DR InterPro; IPR000577; -.
DR InterPro; IPR001452; -.
DR InterPro; IPR001777; -.
DR InterPro; IPR001849; -.
DR InterPro; IPR003006; -.
DR InterPro; IPR003600; -.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00041; fo3; 1.
DR Pfam; PF00047; ig; 48.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RHGEF; 1.
DR PROSITE; PS00933; FGGY_KINASES_1; UNKNOWN_1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
DR SMART; SM00410; IG_Like; 1.
SQ SEQUENCE 6632 AA; 731909 MW; B14BE2EC9AB9C04 CRC64;

Query Match 3.2%; Score 204.5; DB 5; Length 6632;

Best Local Similarity 22.4%; Pred. No. 5.7e-05;

Matches 151; Conservative 98; Mismatches 270; Indels 155; Gaps 31;

Qy 609 EPRYMAVAVHNELIISAIFHTIRFVLASRLOSOMMLMYPATHLHVTVTIGLLIPKFS 668
Db 1238 EPKKEVVVEKQDLSS-----SEVQKETAQQVKEASPEATTITMETSJTSTKT 1286
Qy 669 HSSNNPR-----DDIATEAYEDELDM-----GSGSYLNSINSINAWSE--HSLDPE-DI 714
Db 1287 TTNSTTEVTSTVGVTVEKSESESATTVGGSGGVTEGSISVKIEVVSKTDSQTDV 1346
Qy 715 RDELK---LYAQLIYK-----RKKMTNNPHLOKRCCKGLGRSIMRRITEIPETV 765
Db 1347 REGTPKRVFAEELPVIDSDRKKKKSPDPKKEKSPK-----TEERKASP 1396
Qy 766 SRQCSKEDKAGDHGTAKGTALIRKNPESGNTGSKSETLKNRVFSLKSHSTYDHR 825
Db 1397 TKKTGEVVKSPKSPASPT-----KKEKSPAAEEVKSPTK-----EKSPSS-----P 1440
Qy 826 DOTESSSLPTESQEEETTENSLESGLKLTQKLKEDSEASTESVPLVCKSASHNL 865
Db 1441 TKKEKSPSPTKTGTGDEVKEKS-----PPKSPTKKEKSPKEDVKSPVKKEKSPDATNI 1495
Qy 886 ---SSEKKTGHPTSM---OXSLSVIASAKETGLAGKTQTACGVEERTKSQKPLPK 937
Db 1496 VEVSSSETTIEKTTMTMTMTHESEERTSVKKEKT-----PEKVDKPKS--PTKK 1545

Qy 938 DKETNRNHSNDTETKDPAPQNSNP---AEEPRKPKQSGIMKQQRVNPT-----T 985
Db 1546 DKSPEK-----SITEETKSPVKKEKSPKVEEKPASPTKKEKSPKASPTKKSSENEVKSP 1601
Qy 986 ANSDLNPGTTQMKDNFTDIGEVCWEVYDLTPGVPVPSKVKQKHVSIVASEMEKNPTFSLK 1045
Db 1502 TKKEKSPKSVVEELKSPKESPEKADDPKSPTKKEKSPKS-----ATEDVKSPTKKEK 1657
Qy 1046 --EKSHHKPKAAEVCQOSNOKRID-----KAEVCLWESQGG 1079
Db 1658 SPEKVEEKPTSPTKKESPTKKTDDEVKSPTKKEKSPQTVVEEKPASPTKE-----KSPEK 1713
Qy 1080 SILEDEKLLISKTPVILPERAKEENGQPRAAANCAGOSEELPPKAVA-----SKTENENLN 1135
Db 1714 SVVEEVKSPKES---PEKAAE---KPKSPT---KKEKSPKSAAEVKSPTKKEKSP 1762
Qy 1136 QIGHOEKKTSSSENVGRGYSNNFQOPLTSRAEVCWFEF-ETPAQPNAGRVSVALPASS 1194
Db 1763 EKSAEEKPKSPTKKESPPVKMADEVKSP--TKKEKSPKVEEKPASPT--KKEKTPEKS 1818
Qy 1195 ALSANKIAGPRKEE 1208
Db 1819 A--AEELKSPTKE 1830

Search completed: November 1, 2001, 10:19:21

Job time: 349 sec

OM of: US-09-775-181-2 to: EST:* out_format : pfs

Date: Nov 1, 2001 10:59 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+gpn.model -DEV=xlh  
-Q=/cgn2.1/USPTO_spool/US09775181/runat_30102001_161056_13799/app_query.fasta_1.1955  
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500  
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-ALIGN=15 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-CPU=2000000000 -USER=US09775181@cgn1.1.7084 -NCPU=6  
-XPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY  
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-775-181-2  
Query length: 1215  
Database: EST*  
Database sequences: 10228115  
Database length: 431459454  
Search time (sec): 2318.790000
```

score_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
gb_est31:AU171941	-	663.00	1079.55	5.2e-51	681 ! AU171941 AU171941 Ol-br-ad cdna
gb_gss3:CN0320UA	-	649.00	1052.59	1.2e-49	895 ! AL225307 Tetraodon nigroviridis
gb_gss17:AO837108	-	611.50	996.61	2.2e-46	554 ! AO837108 HS_5464_A1_H02_W7A RPC
gb_gss16:AO764784	-	597.00	973.64	4.1e-45	508 ! AO764784 HS_3118_A1_C10_MR CTP
gb_gss15:AO728799	-	591.00	829.63	4.3e-37	543 ! AO728799 HS_5464_B2_A11_W7A RPC
gb_est63:BB334926	+	438.00	717.02	8.1e-31	280 ! BB334926 BB354926 RIKEN full-le
gb_gss3:CN0322NOT	-	398.00	639.55	1.7e-26	632 ! AL205670 Tetraodon nigroviridis
gb_gss3:CN0322L	-	397.50	633.10	3.8e-26	961 ! AL178348 Tetraodon nigroviridis
gb_gss4:CN04BTK9	-	365.00	578.64	4.1e-23	988 ! AL283590 Tetraodon nigroviridis
gb_gss4:CN04FKT	+	326.50	520.54	7.1e-20	633 ! AL288460 Tetraodon nigroviridis
gb_gss4:CN03PB5	+	326.50	515.51	1.3e-19	921 ! AL254426 Tetraodon nigroviridis
gb_est64:BB429778	+	323.00	525.27	3.9e-20	288 ! BB429778 BB429778 RIKEN full-le
gb_gss4:CN03EM8	-	317.50	499.04	1.1e-18	1030 ! AL240569 Tetraodon nigroviridis
gb_gss34:AZ903235	-	291.50	457.92	2.2e-16	877 ! AL243702 Tetraodon nigroviridis
gb_gss34:AZ903235	-	282.00	450.11	5.9e-16	483 ! AZ903235 RPCT-24-187D3 TJ RPCT
gb_gss4:CN0304C	-	235.50	363.27	4.1e-11	978 ! AL252885 Tetraodon nigroviridis
gb_gss3:CN0320B	+	231.00	356.91	9.2e-11	899 ! AL225308 Tetraodon nigroviridis
gb_est98:BG286130	+	230.50	346.98	3.3e-10	1772 ! BG286130 602383334F1 NIH_MGC.9
gb_gss3:CN0317B	-	229.00	355.03	1.2e-10	807 ! AL223184 Tetraodon nigroviridis
gb_gss3:CN020JP	-	228.00	349.93	2.3e-10	1043 ! AL175678 Tetraodon nigroviridis
gb_gss3:CN02FTA	-	227.50	352.28	1.7e-10	822 ! AL195463 Tetraodon nigroviridis
gb_gss3:CN020UA	-	226.00	347.75	3.0e-10	957 ! AL207163 Tetraodon nigroviridis
gb_gss5:CN05CMB	-	217.00	330.96	2.6e-09	1096 ! AU331292 Tetraodon nigroviridis
gb_est9:AA618743	+	215.00	335.84	1.4e-09	594 ! AA618743 SMOV3MCA1057SK Onchoce
gb_gss4:CN03YL7	-	212.50	326.29	4.7e-09	888 ! AL266452 Tetraodon nigroviridis
gb_gss3:CN035FK	-	210.00	322.87	7.3e-09	840 ! AL228665 Tetraodon nigroviridis
gb_gss5:CN05ATS	+	203.00	311.59	3.1e-08	817 ! AL238969 Tetraodon nigroviridis
gb_gss30:AZ628814	+	201.00	320.45	9.9e-09	329 ! AL262814 1M0481D1F Mouse 10kb
gb_est27:AZ280627	+	200.00	312.71	2.7e-08	518 ! AL280627 4A3A-AAR-F-11-F Anophe
gb_est93:BF859150	+	200.00	299.94	1.4e-07	1343 ! BF859150 963001C06.x4 C. reinh
gb_est82:BF034068	+	197.00	291.48	4.1e-07	1740 ! BF034068 601456406F1 NIH_MGC.6
gb_est100:BG424852	+	196.00	291.90	3.9e-07	1489 ! BG424852 602453107F1 NIH_MGC.1
gb_est86:BF343572	+	195.50	290.83	4.4e-07	1516 ! BF343572 602014403F1 NCI_CGAP
gb_gss4:CN03EM7	-	194.00	306.91	5.6e-08	379 ! AL240568 Tetraodon nigroviridis
gb_gss34:AZ870959	-	193.00	303.06	9.2e-08	446 ! AL2870959 2M0183F08R Mouse 10kb
gb_est89:BF568166	+	190.50	282.20	1.3e-06	1551 ! BF568166 602184033F1 NIH_MGC.4
gb_gss4:CN03Y7G	-	188.50	284.60	9.8e-07	1012 ! BF625957 Tetraodon nigroviridis
gb_est89:BF525803	+	188.50	279.04	2.0e-06	1532 ! BF525803 60206950F1 NCI_CGAP
gb_est80:BF906109	+	186.50	274.49	3.6e-06	1678 ! BF906109 601497158F1 NIH_MGC.7
gb_est91:BF689663	+	185.50	271.82	5.1e-06	1809 ! BF689663 602187063F1 NIH_MGC.4

gb_est95:BG024150 + 183.50 267.66 8.6e-06 1924 ! BG024150 602303029F1 NIH_MG

gb_est86:BF345511 + 181.50 265.07 1.2e-05 1821 ! BF345511 602019270F1 NCI_CG

gb_est99:BG335857 + 181.00 266.79 9.6e-06 1506 ! BG335857 602404570F1 NIH_MG

gb_est85:BF300262 + 181.00 265.49 1.1e-05 1659 ! BF300262 602032170F1 NCI_CG

gb_est99:BF533898 + 180.00 266.23 1.0e-05 1387 ! BF533898 602075261F1 NCI_CG

seq_name: gb_est31:AU171941

seq_documentation_block: 681 bp mRNA EST 29-JAN-2001

LOCUS AU171941

DEFINITION AU171941 Ol-br-ad cdna Oryzias latipes cdna clone br8300, mRNA

SEQUENCE

ACCESSION AU171941

VERSION AU171941

KEYWORDS EST.

SOURCE AU171941.1 GI:12594012

ORGANISM Oryzias latipes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; Belontiiformes; Adrianiichthyidae; Oryziinae; Oryzias.

REFERENCE 1 (bases 1 to 681)

AUTHORS Mita, K., Ishikawa, Y. and Yamauchi, M.

TITLE Establishment of cdna database of medaka, Oryzias latipes

JOURNAL Unpublished (2001)

COMMENT Contact: Mita K

Genome Research Group

National Institute of Radiological Sciences

Anagawa 4-9-1, Inage, Chiba 263-8555, Japan

Email: kmita@nirs.go.jp

method: uni-directional sequence direction: sequenced from T3 primer (5' -> 3')

BASE COUNT 124 a 173 c 185 g 199 t

ORIGIN

Location/Qualifiers

1..681

/organism="Oryzias latipes"

/strain="HNI"

/db_xref="taxon:8090"

/clone_lib="br8300"

/clone_lib="Ol-br-ad cdna"

/sex="female/male mixed"

/tissue_type="brain"

/dev_stage="adult"

Length: 221

Ratio: 3.746

Percent Identity: 63.348

alignment_scores:

Quality: 663.00

Length: 221

Ratio: 3.746

Percent Identity: 63.348

alignment_block:

US-09-775-181-2 x AU171941/rev

Align 'seg 1/1 to reverse of: AU171941 from: 1 to: 681

641 AsptripMetLeuMetLeuTyrPheAlaHisThrHisLeuThrValThrVa 657

681 GACTGTGATGCTCATGCTGTTTGTCTCACATCCTGACTGCTGACTGT 632

657 lThrIleGlyLeuLeuLeulle..ProlypheSerHisSerSerAsnAsn 673

631 GACTCTGGGCTTCTGCTGTTGTTCCCAAGTCTCTGTCAAAAGGATTCAG 582

674 ProArgAspIleAlaThrGluAlaTyGluAspGluLeuAspMetG1 690

591 GCAAGGATGACATCGGCACAGGACATACGAGGAGGCTTGACATGGG 532

690 YArgSerGlySerTyrLeuAsnSerSerIleAsnSerAlaTrpSerGluH 707

531 AAGATCTGGGCTTACCTCAACATACATACATCCTCTGCTGGAGCGAGC 482

707 isSerLeuAspProGluAspIleArgAspGluLeuLysLysLeuTyrAla 723

```

|||||
481 ACAGCTGACCTCATGATATACGGGATGAATTGAAGAGCTGATGCA 432
724 GlnLeuGluIleTyrLysArgLysLysMetIleThrAsnAsnProHisLe 740
|||||
431 CAATCGAAGTCTACAAAGAAAGAGATGCTCGCCACCAACCCACACCT 382
740 uGlnLysLysArgCysSerLysLysGlyLeuGlyArgSerIleMetArgA 757
381 TCAAAAGAAACGCACTCAAGAAAGGCGCTGGGAGCAGCAGCTGATGAAC 332
757 rGileThrGluIleProGluThr.....ValSerArgGlnCysSerLys 771
331 GCATCAGAGATGATCCAGAGACCATCCACCGGCGAGTGCAGCCGA 282
772 GluAspLysGluGlyAlaAspHisGlyThrAlaLysGlyThrAlaLeuI 788
281 GACGAC.....GGCAGCGAGCAGCGCAGCAACCGGGCACC.....TT 244
788 eArgLysAsnProProGluSerSerGlyAsnThrGlyLysSerLysGluG 805
243 GAAGAGAAACCACTTTTGACCCAGC...CACCATGGAAACCTCGGGATG 197
805 luThrLeuLysAsnArgValPheSerLeuLysLysSerHisSerThrTyr 821
196 ACTCCCTTAAAAACAAAGGTCATGGTTTCCGGAAGTCACTCAGC...TTT 150
822 AspHisValArgAspGlnThrGluGluSerSerSerLeuProThrGluSe 838
149 GACCAGTTTGTGACCATGATCCAGAA..... 123
838 rGlnGluGluGluThrGluAsnSerThrLeuGluSerLeuSerGlyL 855
122 .....ACTGGGAACACAGACT...GAGTCTGTTCGGGAG 92
855 ysLysLeuThr 858
|||||
91 ACAAAATGACT 81
seq_name: gb_gss3:CNS032UA

```

```

seq_documentation_block: 895 bp DNA GSS 15-MAY-2000
LOCUS CNS032UA Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 207B04 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL225307
VERSION AL225307.1 GI:7884197
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 895)
AUTHORS Roest-Crollius H., Jaillon O., Dasilva C., Fizes C., Fisher C.,
Bouneau L., Billault A., Quetier F., Saurin W., Bernot A. and
Weissenbach J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 895)
AUTHORS Roest-Crollius H., Jaillon O., Dasilva C., Bouneau L., Fisher C.,
Bernot A., Fizes C., Wincker P., Brottier P., Quetier F.,
Saurin W. and Weissenbach J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 895)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

```

COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES

source Location/Qualifiers
1..895
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone_lib="207B04"
/clone_lib="G"
/notes="Genoscope sequence ID : COAG207DA02LP1-end : T7"

BASE COUNT 107 a 259 c 293 g 236 t
ORIGIN

alignment_scores:

Quality: 649.00 Length: 320
Ratio: 2.785 Caps: 14
Percent Similarity: 72.812 Percent Identity: 51.562

alignment_block:

US-09-775-181-2 x CNS032UA/rev ..

Align seg 1/1 to reverse of: CNS032UA from: 1 to: 895

```

731 LysLysMetIleThrAsnAsnProHisLeu.GlnLysLysArgCysSer. 746
|||||
895 AAGAAGATGCTGGCCACAAACCCACCTCCCCCAAGAAGCGAGCTCCA 846
747 LysLysGlyLeuGlyArgSerIleMetArgIleThrGluIleProGl 763
|||||
845 AAAAAGGGTCTGGCCGCTCGCTGATGAGCGCATCACCGATCCCGGA 796
763 tThrValSerArgGlnCysSerLysGluAspLysGluGlyAlaAspHisG 780
|||||
795 GTCGGTGACCCCAATGCAGCCGCGAGCAAGGAGGCTGGCAGCAGCG 746
780 lyThrAlaLysGly.ThrAlaLeuIleArgLysAsnProProGluSerSe 796
|||||
745 CAACCGCAACAGGTCTGCAGCGTG...AAGAAGAACCCCTTTGATCCAC 699
796 rGlyAsnThrGlyLysSerLysGluGluThrLeuLysAsnArgValPheS 813
|||||
698 CCACGCCCAACCAACCAAGAGGAGGAGGTGTAAAGAGCAAGGTGTTCT 649
813 erLeuLysLysSerHisSerThrTyrAspHisValArgAspGlnThrGlu 829
648 CCCTGAAGAAGTCCACAGCAGCTACGATCAGTCCGTGACCAGGCCGAA 599
830 GluSerSerSerLeuProThrGluSerGlnGluGluThrThrGluAs 846
598 GGGCCCGATGGC.....TCAGCCCGCGACAAAGATGGAGGCTCCGGAAG 555
846 nSerThrLeuGluSerLeuSerGlyLysLysLeuThrGlnLysLeuLysG 863
554 CTCCCTCTCGGAGCGCTCATGGGCAAGAAGATGCCCAAGATCCATCG 505
863 luAspSerGluAla.....GluSerThrGluSerValProLeuValCys 877
504 AAACATCAGTCTCCAGCGAGTCCACCGAGTCCGCTCTCTGCTGCTGC 455
878 LysSerAlaSerAlaHisAsnLeuSerSerGluLysLysThrGlyHisPr 894
454 AAGTCGGCAGCGCTCATAACTCTCGCGGACAAAGAGGCCATTCGCCG 405
894 oArgThrSerMetLeuGlnLysSerSerValIleAlaSerAlaLysG 911
404 CAGAGCTCCATGCTGCAAGAGTCCCTCAGCGTCTATCGCTGGCGCAAG 355
911 luLysThrLeuGlyLeuAlaGlyLys.....ThrGlnThr 922
354 AGAGACCTCGGGTTGACGCCAAAGGCCAGAACATCAGAGGACACCCACA 305

```


AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE
COMMENT

Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu

Plate: 3118 row: E column: 19

Seq primer: M13 Reverse

Class: BAC ends

High quality sequence stop: 508.

FEATURES

source

1. 508

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate-3118 Col-19 Row-E"

/clone_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in

E-Coli DH10B"

BASE COUNT 166 a 129 c 109 g 98 t 6 others

ORIGIN

alignment_scores:

Quality: 597.00 Length: 140
Ratio: 4.489 Gaps: 0
Percent Similarity: 95.000 Percent Identity: 92.143

alignment_block:

US-09-775-181-2 x A0764784 ..

Align seg 1/1 to: A0764784 from: 1 to: 508

715 ArgAspGluLeuLysLysLeuTyrrAlaGlnLeuGluIleTyrrLysArgly 731

88 CAGGACGAGCTGAAAAAATCTCTATGCCAACTGGAATATATAAAGAAA 137

731 slysMetIleThrAsnAsnProHisLeuGlnLysLysArgCysSerLysL 748

138 GAAGATGATCACAAACACCCCTCCAGAAAGCGGCTCGAAGA 187

748 ysGlyLeuGlyArgSerIleMetArgIleThrGluIleProGluThr 764

188 AGGCGCTANGTCGTTCCATCATGACGACCATTCGGAGATCCAGAGACA 237

765 ValSerArgGlnCysSerLysGluAspLysGluGlyAlaAspHisGlyTh 781

238 GTCAGCGCGAGTGTCTAAAGAGACAAAGGAGGCGCGCATGGCAC 287

781 rAlaLysGlyThrAlaLeuLeuArgLysAsnProProGluSerSerGly 798

288 ACCCAAGGCACTGCCCTCATCAGAAAGAACCTCCAGAGTCTTCATGGA 337

798 snThrGlyLysSerLysGluGluThrLeuLysAsnArgValPheSerLeu 814

338 ACACATGGAAATCCCAACGAGGAGACCCCTGAT.AACCGAGTCTTCCTC 386

815 LysLysSerHisSerThrTyrrAspHisValArgAspGlnThrGluGlu 831

387 AAGAATCCCAAGCACTATGACCACTGAGAGACCAACNAGAGATC 436

831 rSerSerLeuProThrGluSerGlnGluGluThrThrGluAsnSerT 848

107 CAAGCAGGCTGAAAAAACTCTATGCCCACTGGATGTATTCAAGAAA 156

437 CAGTACCCTACCCACAGAAAGCCATGAGGAGAGACACAGA.AATTNCA 485

848 hrLeuGluSerLeuSerGly 854

486 CACTGGAATCCNTGTCNGGT 505

seq_name: gb_gss15:A0728799

seq_documentation_block:

LOCUS A0728799 543 bp DNA GSS 15-JUL-1999

DEFINITION HS.5464_B2.All.T7A.RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=1040 Col=22 Row=B, DNA sequence.

ACCESSION A0728799

VERSION A0728799.1 GI:5500351

KEYWORDS GSS.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 543)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)

or from Research h Genetics (info@resgen.com). BAC end Web Server:

http://www.htsc.washington.edu

Plate: 1040 row: B column: 22

Seq primer: T7

Class: BAC ends

High quality sequence stop: 543.

FEATURES

source

1. 543

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate-1040 Col=22 Row=B"

/clone_lib="RPCI-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;

Male blood DNA was isolated from one randomly chosen donor

and partially digested with a combination of EcoRI and

EcoRI Methylase. Size selected DNA was cloned into the

pBACE3.6 vector at EcoRI sites"

BASE COUNT 168 a 139 c 124 g 107 t 5 others

ORIGIN

alignment_scores:

Quality: 511.00 Length: 141
Ratio: 4.223 Gaps: 0
Percent Similarity: 85.816 Percent Identity: 73.050

alignment_block:

US-09-775-181-2 x A0728799 ..

Align seg 1/1 to: A0728799 from: 1 to: 543

715 ArgAspGluLeuLysLysLeuTyrrAlaGlnLeuGluIleTyrrLysArgly 731

107 CAAGCAGGCTGAAAAAACTCTATGCCCACTGGATGTATTCAAGAAA 156

```

731 sLysMetIleThrAsnAsnProHisLeuGlnLysLysArgCysSerLysL 748
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
157 GAGGATGATCTCAACACACCCCTCTCTGATAAAGCGGTGCTCGAAGA 206
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
748 ysGlyLeuGlyArgSerIleMetArgIleThrGluIleProGluThr 764
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
207 AGGCGCTAAGACGTTCCATCATGAGACTCATTACGGAGATCCCAAGACA 256
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
765 ValSerArgGlnCysSerLysGlyAspGlyGluGlyAlaAspHisGlyTh 781
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
257 GACAGACGGCAATGCTCTAAAGAGGACAGAGAGGGCGCGACCATGGCAC 306
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
781 rAlaLysGlyThrAlaLeuLeuArgLysAsnProGluSerSerGlyA 798
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
307 AGTCAAAATGCTCTCTCTCAATCAAGAGAAGACCCCGGAGAGTCTTACGGA 356
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
798 snThrGlyLysSerLysGluGlnThrLeuLysAsnArgValPheSerLeu 814
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
357 ACACACGGAATACAGAGGAGACCTGTAATAACCGAGTCTTCTGCTACTC 406
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
815 LysLysSerHisSerThrTyAspHisValArgAspGlnThrGluGluSe 831
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
407 GAGAAATACACACAGCTCTTATGACCACTGAGACACACACGGAAGACTC 456
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
831 rSerSerLeuProThrGluSerGlnGluGluGluThrThrGluAsnSerT 848
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
457 CTCGTACTACTCTCTCGAAAGCTCAGAGGAGGAGAGCATCATGACTT 506
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
848 hrLeuGluSerLeuSerGlyLys 855
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
507 CACTGTATCTCTCTCAGGAGAA 529
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

seq_name: gb_est63:BB354926

```

seq_documentation_block: 280 bp mRNA EST 12-JUL-2000
LOCUS BB354926
DEFINITION BB354926 RIKEN full-length enriched, adult male corpus striatum Mus
            musculus cDNA clone C030003G02 3', mRNA sequence.
ACCESSION BB354926
VERSION BB354926.1 GI:9066754
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

```

REFERENCE
AUTHORS
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, X., Watanabe, S., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshinori Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL: http://genome.rtc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE
JOURNAL
COMMENT

Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

Source
1. .280
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C030003G02"
/clone_lib="RIKEN full-length enriched, adult male corpus striatum"
/sex="male"
/tissue_type="corpus striatum"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
GAGAGAGAGAGGATCCAGAGACTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTAAATAATATCCCGCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 79 a 65 c 62 g 74 t
ORIGIN

alignment_scores:
Quality: 438.00 Length: 93
Ratio: 4.867 Gaps: 0
Percent similarity: 96.774 Percent Identity: 91.398
alignment_block:
US-09-775-181-2 x BB354926
Align seg 1/1 to: BB354926 from: 1 to: 280
633 ValLeuAlaSerArgLeuGlnSerAspTrpMetLeuMetLeuTyrPheAl 649
1 GTGCTAGCTTCGAGATTTCACCGCGATGGATCTGCTGCTACTTGC 50
649 aHisThrHisLeuThrValThrValThrIleGlyLeuLeuLeuProL 666
51 ACATGCTTACTTGACTGTGACAGTTACCACTGGGCTGCTTAAATCCA 100
566 ysPheSerHisSerSerAsnAsnProArgAspAspIleAlaThrGluAla 682
101 AGTTTTCACATTCAGCAATTAATCCCGGAGAGACATTTGCTACGAAGA 150
683 TyrGluAspGluLeuAspMetGlyArgSerGlySerTyrLeuAsnSerSe 699
151 TATGAAGATGAGCTGGACATGGTTCGCTCTGCGATCCTACCTGAACAGCAG 200


```

667 .....PheSerHi 669
303 CCGTTCCGGGTCTGAAACCGTATCTGTGCCCTCCACCGATTCTCTCTT 352
669 sSerSerAsnAsnProArgAspAlaThrGluAlaThrGluAspG 686
353 TGCCGGCACCACATCGGACGACATGGCTCGGAGGCTACGAGACG 402
686 luleuAspMetGlyArgSerGlySerTyrLeuAsnSerSerIleAsnSer 702
403 AGCTGGACATGGGGCGCTCGGGTCTTACCTCAACAGCAGCATCACGTCG 452
703 AlaTrpSerGluHisSerLeuAspProGluAspIleArg 715
453 CGTGGAGCGGACGACACCGCTGGATCCCGAAGACATTCGG 491

seq_name: gb_gss4:CNS03PB5

seq_documentation_block:
LOCUS CNS03PB5 921 bp DNA GSS 17-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
044B20 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL254426
VERSION AL254426.1 GI:7975438
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 921)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 921)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 921)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
Location/Qualifiers
1..921
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="044B20"
/clone_lib="G"
/note="Genoscope sequence ID : C08C044DA10SPI-end :
PUC-ori"
BASE COUNT 200 a 281 c 248 g 192 t
ORIGIN

alignment_scores:
Quality: 326.50 Length: 113
Ratio: 3.934 Gaps: 1
Percent Similarity: 73.451 Percent Identity: 57.522

alignment_block:
US-09-775-181-2 x CNS03PB5

```

Align seg 1/1 to: CNS03PB5 from: 1 to: 921

```

626 IlePheHisThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTr 642
:::|||||
478 CTGTTCTCCTCGACAGGTTCACTCTGTGTCAGGAGCTCCACCCAGACTG 527
612 pMetLeuMetLeuTyrPheAlaHisThrHisLeuThrValThrI 659
|||||
528 GATGCTGCTGCTGTTCATCCACACACACACCTCACCGTCACCGTCACGC 577
659 leGlyLeuLeuLeuLeuLeuProLys..... 666
:::|||||
578 TGGCGCTGCTGCTGCTGCCCAAGGTGAGCGTCCGCGTAGCTCTCTGTCCA 627
667 .....PheSerHi 669
|||||
628 CCGGTTCCGGGTCTCTGAAACCGTATCTGTGCCCTCCACCGATTCTCTCTT 677
669 sSerSerAsnAsnProArgAspAlaThrGluAlaThrGluAspG 686
:::|||||
678 TGCCGGCACCACATCGGACGACATGGCTCGGAGGCTACGAGACG 727
686 luleuAspMetGlyArgSerGlySerTyrLeuAsnSerSerIleAsnSer 702
:::|||||
728 AGCTGGACATGGGGCGCTCGGGTCTTACCTCAACAGCAGCATCACGTCG 777
703 AlaTrpSerGluHisSerLeuAspProGluAspIleArg 715
|||||
778 CGCTGGAGCGGACGACACCGCTGGATCCCGAAGACATTCGG 816

```

seq_name: gb_est64:BB429778

```

seq_documentation_block:
LOCUS BB429778 288 bp mRNA EST 18-JUL-2000
DEFINITION BB429778 RIKEN full-length enriched, adult male hippocampus Mus
musculus cDNA clone C630030A14 3', mRNA sequence.
ACCESSION BB429778
VERSION BB429778.1 GI:9269505
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 288)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,
M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL: http://genome.rtc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length

```

t.
47

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 10:14:19 ; Search time 45:13 Seconds
(without alignments)
804.648 Million cell updates/sec

Title: US-09-775-181-4

Perfect score: 3199

Sequence: 1 MGAWAYPELLCLLQAQLGLG.....YMTAVGMWSLVSDGLTIFQ 599

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

al number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601.*
1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	141.5	4.4	974	21	AA45000 Human calcium sens
2	139.5	4.4	917	20	AA49126 Chimeric hCAR/hmgl
3	139.5	4.4	1276	20	AA49127 PhCAR/hmglur2*Gai5
4	136	4.3	901	21	AA45001 Human calcium sens
5	133.5	4.2	1026	18	AAW32059 Dogfish shark kidn
6	133.5	4.2	1079	19	AAW54847 Rat kidney calcium
7	133.5	4.2	1079	19	AAW38275 Rat kidney cell ca
8	133.5	4.2	1079	20	AA41781 Rat parathyroid ca
9	133.5	4.2	1079	20	AAW4928 Rat kidney extrace
10	133.5	4.2	1079	20	AAW89566 Rat parathyroid ca
11	133.5	4.2	1079	21	AA451828 Rat calcium recept

12	132	4.1	1001	21	AA44999 Human calcium sens
13	127.5	4.0	840	21	AA45021 Rat sensory transd
14	127.5	4.0	840	21	AA45026 Rat sensory transd
15	127.5	4.0	840	21	AA45027 Rat sensory transd
16	127.5	4.0	840	21	AA45028 Rat sensory transd
17	127	4.0	1078	17	AAW11889 Parathyroid calciu
18	127	4.0	1078	19	AAW54846 Human parathyroid
19	127	4.0	1078	19	AAW38274 Human parathyroid
20	127	4.0	1078	20	AA428840 Human calcium rece
21	127	4.0	1078	20	AA41780 Human parathyroid
22	127	4.0	1078	20	AAW89565 Human parathyroid
23	127	4.0	1078	21	AA451827 Human calcium rece
24	127	4.0	1078	21	AA470325 Human wild type ca
25	123	3.8	1085	19	AAW54844 Bovine parathyroid
26	123	3.8	1085	19	AAW38272 Bovine parathyroid
27	123	3.8	1085	20	AA41778 Bovine parathyroid
28	123	3.8	1085	20	AAW89563 Bovine parathyroid
29	123	3.8	1085	21	AA451825 Bovine calcium rec
30	120	3.8	824	20	AA49124 Human mglur2 prote
31	120	3.8	872	17	AAW89580 Metabotropic gluta
32	120	3.8	1035	20	AA49128 Chimeric hmglur2/h
33	120	3.8	1394	20	AA49129 pmglur2/CaR*Galpha
34	120	3.8	1397	20	AA49134 pmglur2/CaR*Galpha
35	119.5	3.7	1088	17	AAW11888 Parathyroid calciu
36	119.5	3.7	1088	19	AAW54845 Human parathyroid
37	119.5	3.7	1088	19	AAW38273 Human parathyroid
38	119.5	3.7	1088	20	AA41779 Human parathyroid
39	119.5	3.7	1088	20	AAW89564 Human parathyroid
40	119.5	3.7	1088	21	AA451826 Human calcium rece
41	118.5	3.7	1058	18	AAW25762 Amino acid sequenc
42	118	3.7	872	17	AAW95052 Human metabotropic
43	116	3.6	906	16	AAW64250 Human mglur1b. Ho
44	116	3.6	1056	18	AAW25765 Amino acid sequenc
45	116	3.6	1194	14	AAW42199 HSmglur1. Homo sa

ALIGNMENTS

RESULT 1
AA45000
ID AA45000 standard; Protein; 974 AA.

XX AA45000;

XX 20-JUN-2000 (first entry)

XX Human calcium sensing receptor isoform, CaSRd.

XX Calcium Sensing Receptor; CaSRd; isoform; human; splice variant;
KW hypotensive; osteopathic; receptor activity; calcium level; modulator;
KW treatment; hyperparathyroidism; osteoporosis; Paget's disease;
KW hypercalcaemia malignancy; hypertension; gene therapy.

XX Homo sapiens.

OS Synthetic.

XX WO200006601-A1.

XX 10-FEB-2000.

XX 28-JUL-1999; 99WO-US17116.

XX 30-JUL-1998; 98US-0094702.

XX (AVET) AVENTIS PHARM PROD INC.

XX Yu KT, Labaudiniere RF, Thrower LW;

XX WPI: 2000-195263/17.

XX N-PSDB; AA50616.

XX Nucleic acids encoding isoforms of human calcium sensing receptor for

QY 397 EG-----CPFCADD-----SPCFVQEDKYLK-----LAIISQGLCMLLDFV 433
Db 577 egefsdendasactkcpndfwsnhtsciakeyleylswtepfialtifavilgilitf 636
QY 434 SMLVVVHFRKAKSTRASGLILLETILFSGSLLYFPVVLVFEPTFCILLRWARLLGFA 493
Db 637 vlgvfikrntpvkatnrslsylllflccfssllfgeprdwtrlrqafglsfv 696
QY 494 TVYGTVTLLKHLRWLVFLSRTAQRIPTYMTGGRVVRMLAVILV-----VFVFLIGWTS 546
Db 697 lciscilvktnrllvfeakiptslhrkwwglnqlfllvflcivqlvtocliwlytapps 756
QY 547 SVCQNLKQISLIGQKTSDFHLENCLDRWDYMTAVGWSLSVSY 592
Db 757 sy-rnhele-----deviifitc-----degslmaigfligy 786

RESULT 6

AAW54847
AAW54847 standard; Protein; 1079 AA.

AAW54847;

01-SEP-1998 (first entry)

Rat kidney calcium receptor 3A protein 4Kb fragment.

Calcium ion concentration; parathyroid hormone; homeostasis; kidney;
Calcium receptor; detection.

Rattus sp.

US5763569-A.

09-JUN-1998.

07-JUN-1995; 95US-0484565.

07-JUN-1995; 95US-0484565.

23-AUG-1991; 91US-0749451.

11-FEB-1992; 92US-0834044.

21-AUG-1992; 92US-0934161.

12-FEB-1993; 93US-0017127.

23-FEB-1993; 93US-0009389.

22-OCT-1993; 93US-0141248.

19-AUG-1994; 94US-0292827.

21-OCT-1994; 94WO-0512117.

08-DEC-1994; 94US-0353784.

(BGHM) BRIGHAM & WOMENS HOSPITAL.

(NPSP-) NPS PHARM INC.

Brown EM, Garrett JE, Hebert SC;

WPI; 1998-347412/30.

N-PSDB; AAV26965.

Calcium receptor poly:peptide(s) - useful for drug screening or
antibody production

Claim 8; Fig 50; 174pp; English.

The tissue from which the rat kidney calcium receptor and receptors from
bovine parathyroid and rat kidney are derived, respond to changes, and
control changes, in calcium ion concentration, e.g. parathyroid hormone
regulates Ca²⁺ homeostasis in blood and extracellular fluid, and kidney
function alters through changes in Ca²⁺ levels in juxtaglomerular and
proximal tubule cells in the kidney. The purified receptors (produced
recombinantly) can be used to screen for compounds that modulate calcium
receptor activity, especially those that can be used to treat diseases
associated with the receptors in these tissues. They can also be used
to raise antibodies for use in detection assays.

XX

XX
SQ Sequence 1079 AA;

Query Match 4.2%; Score 133.5; DB 19; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.0016;
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;

QY 235 HRRGPNQGRGLGHSRRKDKLGDSKSHFKWSPPYLECEGSKYKPGWLVTLSAIYGLQP 294
Db 377 heeg---gnrlsstafprlctgdeninsvetpymdyehlrisyvnylavysiahald 433

QY 295 --NLVFEFRGVMKVDINLQKVIDQCSSDGFSGTHKCHLNNSS-----ECMPIKG 342
Db 434 iytclpg-rglf---tngscadikkv--eawqvlklrhlrlnftnmngeqvtfdccgdllg 487

QY 343 LGFVL-----GAYECICKAG---FVHPGVLPVNNFRRRPGDOHIS--- 379
Db 488 nysiinwhlspedgsvfkevgyynyvaykgerlfinneekilwsgfsrevpfnsrddc 547

QY 380 -GSTKDVSEE---AYVCLPCREG-----CPFCADD-----SPCFVQEDKYLK 417
Db 548 agtrkgliieptccfecvcpdgyesgetdascdkcpddfwsnhtsciakeiefla 607

QY 418 -----LAIISFOGLCMLLDFVSMVLVYHFRKAKSTRASGLILLETILFSGSLLYFPVVI 471
Db 608 wtepfialtlfavlglftafvlgvfikrntpvkatnrslsylllflccfssslf 667

QY 472 LYFEPSTFCILLRWARLLGFAIVYGTVTLLKHLRWLVFLSRTAQRIPTYMTGGRVVRMLA 531
Db 668 figepqdwtrlrqafglsfviciscilvktorvlvfeakiptsfhrkwwglnqlfll 727

QY 532 V-----ILLVFWFLIGWTSVVCQNLKQISLIGQKTSDFHLENCLDRWDYMTAV 584
Db 728 vflctfmqliicliwlytappssy-rnhele-----deiifitc---hegsimal 773

QY 585 GMSLSVSY 592

Db 774 g---sligy 779

RESULT 7

AAW38275
ID AAW38275 standard; Protein; 1079 AA.

AC AAW38275;

DT 08-MAY-1998 (first entry)

Rat kidney cell calcium receptor 3A (RakCar 3A).

Rat kidney cell calcium receptor 3A; RakCar 3A;

calcium homeostasis; hyperparathyroidism; osteoporosis.

Rattus sp.

US5688938-A.

18-NOV-1997.

07-JUN-1995; 95US-0485588.

07-JUN-1995; 95US-0485588.

23-AUG-1991; 91US-0749451.

11-FEB-1992; 92US-0834044.

21-AUG-1992; 92US-0934161.

12-FEB-1993; 93US-0017127.

23-FEB-1993; 93US-0009389.

22-OCT-1993; 93US-0141248.

19-AUG-1994; 94US-0292827.

21-OCT-1994; 94WO-0512117.

08-DEC-1994; 94US-0353784.

XX

Db 608 wtepfqialtlfavlgltavlgvfkfrntpvkatnrelsylllfsllccfsslf 667
 QY 472 LYEPSTFRFCILLRWARLLGFATVYGVTKLHRVLKRVLSRTAQRIPIYWTGGRVWRMLA 531
 Db 668 fipepdwtcrlrqpafglsfvlciscilvktnrllvfeakiptsfhrkwwglnlqfll 727
 QY 532 V-----ILLVVFELIGWTSSVCQNLKQISLIGGKTSDDLIFNMCLIDRWDMYTA 584
 Db 728 vflctfmqlliciwiyltappssy-rnhele-----deifitc---hegsimal 773
 QY 585 GMWSLSVSY 592
 Db 774 g--sligy 779

RESULT 9
 AAW94928
 ID AAW94928 standard; Protein; 1079 AA.
 XX AAW94928;
 04-MAY-1999 (first entry)
 Rat kidney extracellular calcium/polyvalent cation-sensing receptor.
 DE DE
 XX Phormone receptor; signal transduction; fertility; behaviour;
 XX reproduction; rodent; insect; rat.
 OS Rattus sp.
 XX WO9900422-AL.
 XX 07-JAN-1999.
 XX 30-JUN-1998; 98WO-US13680.
 XX 30-JUN-1997; 97US-0051284.
 XX (HARD) HARVARD COLLEGE.
 PA Buck L, Dulac C, Herrada G, Matsunami H;
 PI WPI; 1999-095684/08.
 XX New isolated phormone receptor polypeptides - used to develop
 PT products for controlling fertility and behaviour in vertebrates and
 PT invertebrates
 Disclosure; Page 169-170; 308pp; English.

CC The invention relates to polynucleotide sequences encoding mammalian
 CC phormone receptor polypeptides. The polypeptides are expressed in
 CC murine and rat vomeronasal organ. The products can be used for modifying
 CC phormone activity, e.g. for decreasing phormone receptor mediated
 CC signal transduction. They can be used for controlling fertility and
 CC behaviour in vertebrates and invertebrates. Compositions comprising the
 CC polypeptides are particularly useful in e.g. controlling fertility in
 CC livestock and controlling reproduction in rodents or insects by
 CC interrupting the normal behaviours of rodents or insects that result in
 CC reproduction. The present sequence represents a rat kidney extracellular
 CC calcium/polyvalent cation-sensing receptor.
 XX Sequence 1079 AA;

Query Match 4.2%; Score 133.5; DB 20; Length 1079;
 Best Local Similarity 21.7%; Pred. No. 0.0016;
 Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;

QY 235 HRRGPNQGRGLGHSWRRKDGGLGDKSHFKWSPPYLECENGSYKPGWLVTLSSAIYGLQP 294
 Db 377 heeg---gnrlnsstafplcldgeninsvetpymdyehlrlnsvnylavysiahald 433

QY 295 --NLVPEFRGVMKVDINLQKVIDIDQCSSDGWFSGTHKCHLNN-----ECMPIKG 342
 Db 434 iytclp-rglf---tngscadlkv--eawqvlklrhlnfnmngseqvtfdcdglvg 487
 QY 343 LGFVL-----GAYECICRAG---FYHPGVLPVNNFRRRPGDQHS---- 379
 Db 488 nyslinwhlspedgsivfkevyyynyakkgerlfineekilwsgfsrevpfncardcq 547
 QY 380 -GSTKDVSE-----AVCLPCREG-----CPFCADD-----SPCFVQEDKYL 417
 Db 548 agtrkgliegeptccfecvecpdgyesgetdasackcpddfwsnenhtsciakiefia 607
 QY 418 -----LAISFOGLCMLLDFVSMVYVYHFRKAKSRASCLILLETILFSGSLLLYFPVVI 471
 Db 608 wtepfqialtlfavlgltavlgvfkfrntpvkatnrelsylllfsllccfsslf 667
 QY 472 LYEPSTFRFCILLRWARLLGFATVYGVTKLHRVLKRVLSRTAQRIPIYWTGGRVWRMLA 531
 Db 668 fipepdwtcrlrqpafglsfvlciscilvktnrllvfeakiptsfhrkwwglnlqfll 727
 QY 532 V-----ILLVVFELIGWTSSVCQNLKQISLIGGKTSDDLIFNMCLIDRWDMYTA 584
 Db 728 vflctfmqlliciwiyltappssy-rnhele-----deifitc---hegsimal 773
 QY 585 GMWSLSVSY 592
 Db 774 g--sligy 779

RESULT 10
 AAW89566
 ID AAW89566 standard; Protein; 1079 AA.
 XX AAW89566;
 XX 19-MAR-1999 (first entry)
 XX Rat parathyroid calcium receptor pRakCar 3A.
 DE DE
 KW Parathyroid calcium receptor; inorganic ion receptor; osteoporosis;
 KW calcium homeostasis; hyperparathyroidism; seizure; stroke; epilepsy;
 KW spinal cord injury; hypoxia-induced nerve cell damage; cardiac arrest;
 KW neonatal distress; neurodegenerative disease; Alzheimer's disease;
 KW Huntington's disease; Parkinson's disease; dementia; muscle tension;
 KW depression; anxiety.
 OS Rattus sp.
 XX US5858684-A.
 XX 12-JAN-1999.
 XX 07-JUN-1995; 95US-0480751.
 XX 07-JUN-1995; 95US-0480751.
 PR 23-AUG-1991; 91US-0749451.
 PR 11-FEB-1992; 92US-0834044.
 PR 21-AUG-1992; 92US-0934161.
 PR 12-FEB-1993; 93US-0017127.
 PR 23-FEB-1993; 93US-0009389.
 PR 22-OCT-1993; 93US-0141248.
 PR 19-AUG-1994; 94US-0292827.
 PR 21-OCT-1994; 94WO-US12117.
 PR 08-DEC-1994; 94US-0353784.
 XX (BGM) BRIGHAM & WOMENS HOSPITAL.
 PA (NPSF-) NPS PHARM INC.
 XX Balandrin MF, Brown EM, Del Mar EG, Garrett JE;
 PI Hebert SC, Nemeth EF, Van Wagenen BC;
 XX WPI; 1999-119871/10.
 DR N-PSDB; AAW82486.

AA45021
ID AAY45021 standard; Protein; 840 AA.
XX AC AAY45021;
XX DT 31-MAY-2000 (first entry)
XX DE Rat sensory transduction G-protein coupled receptor-B3.
XX KW Rat; sensory transduction; G-protein coupled receptor-B3; GPCR-B3;
XX KW sensory cell; taste receptor cell; screen; taste modulator;
XX KW pharmaceutical; food industry; taste topographic map; tongue.
XX OS Rattus sp.
XX FH Key Location/Qualifiers
XX FT Domain 1..580
XX FT /label= Extracellular_domain
XX PN WO200006592-A1.
XX 10-FEB-2000.
XX PF 27-JUL-1999; 99WO-US17099.
XX PR 28-JUL-1998; 98US-0094465.
XX PA (REGC) UNIV CALIFORNIA.
XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX PI Zuker CS, Adler JE, Lindemeier J, Ryba N, Hoon M;
XX WPI; 2000-205451/18.
XX DR N-PSDB; AAZ50743.
XX PT New isolated sensory transduction G-protein coupled receptor, useful
XX PT for developing products for use in studying and modulating the taste
XX PT transduction pathway -
XX PS Claim 22; Page 75; 83pp; English.
XX CC The present sequence is a taste cell specific G-protein
XX CC coupled receptor, GPCR-B3 which is involved in sensory transduction.
XX CC This sequence was isolated from the IZAP rat circumvallate cDNA library.
XX CC GPCR-B3 is specifically expressed in foliate and fungiform cells, with
XX CC lower expression in circumvallate taste receptor cells of the tongue.
XX CC The protein has an extracellular domain, seven transmembrane
XX CC domains and an intracellular domain.
XX CC The present sequence is used to screen compounds that modulate sensory
XX CC signalling in taste cells, especially taste modulators useful in
XX CC pharmaceutical and food industries to customise taste. The sequence
XX CC can also be used as probe for identifying taste cells and
XX CC subsets of taste receptor cells such as foliate, fungiform and
XX CC circumvallate. Such probes are also useful to generate taste
XX CC topographic maps that elucidate the relationship between the taste
XX CC cells of the tongue and sensory neurons leading to taste centres
XX CC in the brain.
XX SQ Sequence 840 AA;
Query Match 4.08; Score 127.5; DB 21; Length 840;
Best Local Similarity 22.24; Pred. No. 0.0041;
Matches 60; Conservative 48; Mismatches 89; Indels 73; Gaps 14;
QY 306 VDINLQK-----VDIDQSSDGFSCGTHKCHLANSCEMPKIGLGVLAGVECI-CK 355
DB 475 idinktiqhwkngqvpvsctcd-claghrhvvvghhc-----cfcvpcpe 522
QY 356 AGFYHGPVLPVNNRRRGPQDHISGTSKDVSEAYVCLPC--REGCPFCADDSPCFVQED 413
DB 523 agtf-----ln-----mselhicqpcgteawp--kesttcfrtv 556

QY 414 KYL-----RLAIIISFOGLCMLDGF-VSMLVVYHFRKAKSTRASG-----LILLETILFGS 462
DB 557 eflawhepisiiviaantllllllivtagllfawhfh-tpvvrsgaggricflmglsvlavs 615
QY 463 LLLYFPVVILYFEPSTFRCILLRWARLLGFATVYGVTVLKLHRLVKVFLSRTAQRIPLYMT 522
DB 516 csfy-----sffgeptvpacilrqlfsgfaifscilrfsqvlifkftskvptfyr 671
QY 523 -----GGRVMRLAVILLVVFVFLIGWT 545
DB 672 waqhggaglfvsvstvhllcltvlwmwt 701
RESULT 14
AAY45026
ID AAY45026 standard; Protein; 840 AA.
XX AC AAY45026;
XX DT 31-MAY-2000 (first entry)
XX DE Rat sensory transduction G-protein coupled receptor-B3 variant #1.
XX KW Rat; sensory transduction; G-protein coupled receptor-B3; GPCR-B3;
XX KW sensory cell; taste receptor cell; screen; taste modulator;
XX KW pharmaceutical; food industry; taste topographic map; tongue; variant.
XX OS Rattus sp.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT Misc-difference 33
XX FT /note= "Wild type Leu substituted by Ile"
XX PN WO200006592-A1.
XX PD 10-FEB-2000.
XX PF 27-JUL-1999; 99WO-US17099.
XX PR 28-JUL-1998; 98US-0094465.
XX PA (REGC) UNIV CALIFORNIA.
XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX PI Zuker CS, Adler JE, Lindemeier J, Ryba N, Hoon M;
XX WPI; 2000-205451/18.
XX PT New isolated sensory transduction G-protein coupled receptor, useful
XX PT for developing products for use in studying and modulating the taste
XX PT transduction pathway -
XX PS Disclosure; Page -; 83pp; English.
XX CC The present sequence is a polymorphic variant #1 of rat taste cell
XX CC specific G-protein coupled receptor, GPCR-B3 which is involved in
XX CC sensory transduction. GPCR-B3 is specifically expressed in foliate
XX CC and fungiform cells, with lower expression in circumvallate taste
XX CC receptor cells of the tongue. The protein has an extracellular domain,
XX CC seven transmembrane domains and an intracellular domain.
XX CC The GPCR-B3 sequence is used to screen compounds that modulate sensory
XX CC signalling in taste cells, especially taste modulators useful in
XX CC pharmaceutical and food industries to customise taste. The sequence
XX CC can also be used as probe for identifying taste cells and
XX CC subsets of taste receptor cells such as foliate, fungiform and
XX CC circumvallate. Such probes are also useful to generate taste
XX CC topographic maps that elucidate the relationship between the taste
XX CC cells of the tongue and sensory neurons leading to taste centres
XX CC in the brain.
XX CC Note: The present sequence is not given in the specification but is
XX CC derived from rat GPCR-B3 sequence shown in page 75 (AAY45021).

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 10:13:25 ; Search time 26.44 Seconds
(without alignments)
466.476 Million cell updates/sec

Title: US-09-775-181-4
Perfect score: 3199
Sequence: 1 MGNMAYPLLCLLLAQLGLG.....YMTAVGMWSLVSYDGLTIFQ 599

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.AA.*
1: /cgn2.6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2.6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2.6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2.6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2.6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2.6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	133.5	4.2	1079	1 US-08-485-588-8	Sequence 8, Appli
2	133.5	4.2	1079	1 US-08-484-565-8	Sequence 8, Appli
3	133.5	4.2	1079	2 US-08-480-751-8	Sequence 8, Appli
4	133.5	4.2	1079	3 US-08-943-986-8	Sequence 8, Appli
5	133.5	4.2	1079	3 US-08-353-784-8	Sequence 8, Appli
6	133.5	4.2	1079	3 US-08-484-719B-8	Sequence 8, Appli
7	127	4.0	1078	1 US-08-485-588-7	Sequence 7, Appli
8	127	4.0	1078	1 US-08-484-565-7	Sequence 7, Appli
9	127	4.0	1078	2 US-08-480-751-7	Sequence 7, Appli
10	127	4.0	1078	2 US-08-943-986-7	Sequence 7, Appli
11	127	4.0	1078	3 US-08-353-784-7	Sequence 7, Appli
12	127	4.0	1078	3 US-08-484-719B-7	Sequence 7, Appli
13	123	3.8	1085	1 US-08-485-588-5	Sequence 5, Appli
14	123	3.8	1085	1 US-08-484-565-5	Sequence 5, Appli
15	123	3.8	1085	2 US-08-480-751-5	Sequence 5, Appli
16	123	3.8	1085	2 US-08-943-986-5	Sequence 5, Appli
17	123	3.8	1085	3 US-08-353-784-5	Sequence 5, Appli
18	123	3.8	1085	3 US-08-484-719B-5	Sequence 5, Appli
19	121.5	3.8	1059	4 US-09-134-513-2	Sequence 2, Appli
20	119.5	3.7	1088	1 US-08-485-588-6	Sequence 6, Appli
21	119.5	3.7	1088	1 US-08-484-565-6	Sequence 6, Appli
22	119.5	3.7	1088	2 US-08-480-751-6	Sequence 6, Appli
23	119.5	3.7	1088	2 US-08-943-986-6	Sequence 6, Appli
24	119.5	3.7	1088	3 US-08-353-784-6	Sequence 6, Appli
25	119.5	3.7	1088	3 US-08-484-719B-6	Sequence 6, Appli
26	118.5	3.7	1058	2 US-08-687-289A-5	Sequence 5, Appli
27	118	3.7	872	3 US-08-337-797A-2	Sequence 2, Appli

28	118	3.7	872	3 US-09-258-523-2	Sequence 2, Appli
29	116	3.6	906	1 US-08-486-270-2	Sequence 2, Appli
30	116	3.6	906	1 US-08-367-264-2	Sequence 2, Appli
31	116	3.6	1056	2 US-08-687-289A-8	Sequence 8, Appli
32	115	3.6	906	5 PCT-US91-09422-17	Sequence 17, Appli
33	115	3.6	1199	1 US-08-041-538-2	Sequence 2, Appli
34	115	3.6	1199	1 US-08-463-642-2	Sequence 2, Appli
35	115	3.6	1199	1 US-08-455-602-2	Sequence 2, Appli
36	115	3.6	1199	2 US-08-465-157-2	Sequence 2, Appli
37	115	3.6	1199	5 PCT-US91-09422-2	Sequence 2, Appli
38	114.5	3.6	879	1 US-08-072-574-6	Sequence 6, Appli
39	114.5	3.6	879	1 US-08-486-270-6	Sequence 6, Appli
40	114.5	3.6	879	3 US-08-367-264-6	Sequence 6, Appli
41	114.5	3.6	1219	2 US-08-687-289A-6	Sequence 6, Appli
42	114	3.6	1180	4 US-08-660-148-2	Sequence 2, Appli
43	114	3.6	1212	4 US-08-660-148-5	Sequence 5, Appli
44	113	3.5	877	1 US-08-072-574-12	Sequence 12, Appli
45	113	3.5	877	1 US-08-486-270-12	Sequence 12, Appli

ALIGNMENTS

RESULT 1
US-08-485-588-8
; Sequence 8, Application US/08485588
; Patent No. 5688938
; GENERAL INFORMATION:
; APPLICANT: Edward M. Brown
; APPLICANT: Steven C. Hebert
; APPLICANT: Forrest H. Fuller
; APPLICANT: James E. Garrett, Jr.
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; MOLECULES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTSEQ
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,588
; FILING DATE: 7 June, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 9
; APPLICATION NUMBER: 08/353,784
; FILING DATE: 9 December, 1994
; APPLICATION NUMBER: PCT/US/94/12117
; FILING DATE: 21 October, 1994
; APPLICATION NUMBER: U.S. 08/292,827
; FILING DATE: 23 August, 1994
; APPLICATION NUMBER: U.S. 08/141,248
; FILING DATE: 22 October, 1993
; APPLICATION NUMBER: U.S. 08/009,389
; FILING DATE: 23 February, 1993
; APPLICATION NUMBER: U.S. 08/017,127
; FILING DATE: 12 February, 1993
; APPLICATION NUMBER: U.S. 07/934,161
; FILING DATE: 21 August, 1992
; APPLICATION NUMBER: U.S. 07/834,044
; FILING DATE: 11 February, 1992
; APPLICATION NUMBER: U.S. 07/749,451

;; FILING DATE: 23 August, 1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Heber, Sheldon O.
;; REGISTRATION NUMBER: 38,179
;; REFERENCE/DOCKET NUMBER: 213/005
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (213) 489-1600
;; TELEFAX: (213) 489-0440
;; TELEX: 67-3510
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1079 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-485-588-8

Query Match 4.2%; Score 133.5; DB 1; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.00041;
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;
QY 235 HRRGPNQGRGLGHSWRKDGKSHFKWSPYLECENGSKPGWLVTLSAIYGLQP 294
DB 377 HEEG---GNRLNSSTAPRPLCTGDNINSVETPYMDYEHRLRISYNYLAVYSIAHALQD 433
QY 295 --NLVPERGVKVDINLQKVDIDQCSSDGSFSGTHKHLNNS-----ECMPIKG 342
DB 434 IYTCPLPG-RGLF---TNGSCADIKV--EAMQVLKHLRLHNFNTNMGEQVTFDECGDLVG 487
QY 343 LGFVL-----GAYECICKAG---FYHPGVLPVNNRRRGGPDQHS-----SPCFVQEDKYL 417
DB 488 NYSIINHLSPEDGSIVFKEVGYNYVAKKGERLFINEEKILWGSFREVFPNSCRDQC 547
QY 380 -GSTKDVSEE---AYVCLPCREG-----CPFCADD-----SPCFVQEDKYL 417
DB 548 AGTRKGIIEGPTCCFCEVCEPDGYSGETDASACDKCPDFFWSNENHTSCIAKEIEFLA 607
QY 418 -----LAIISFQGLCMMLDFVSMVYHFRKAKSIRASGLILLETILFGSILLYFPVVI 471
DB 608 WTEPFGLTALFAVLGIFLTAFLVGVFIKFRNTPIVKAATNRELSYLLFLSLCCFSSSLF 667
QY 472 LYFEPSTRCILLRWALLGATVYGTWTLKHLRVKVLFSRTAQRIPTYMTGGRVMRLA 531
DB 668 FIGEPQDWTCLRQPAFGISVLCISCLVKTNRVLLVFEAKIPTSFHRKWWGLNQLQFL 727
QY 532 V-----ILLVWFELIGWTSVYCONLEKQISLIGOGKTSDDLIFNMCLIDRDWYMTAV 584
DB 728 VFLCTFMQLLICIWLTYTAPSSY-RNHELE-----DELIIFTTC---HEGSLMAL 773
QY 585 GMSLSVSY 592
DB 774 G--SLIGY 779

RESULT 2
US-08-484-565-8
; Sequence 8, Application US/08484565
; Patent No. 5763569
; GENERAL INFORMATION:
; APPLICANT: Edward M. Brown
; APPLICANT: Steven C. Hebert
; APPLICANT: James E. Garrett, Jr.
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California

COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,565
FILING DATE: 7 June, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 9
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1079 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-565-8

Query Match 4.2%; Score 133.5; DB 1; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.00041;
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;
QY 235 HRRGPNQGRGLGHSWRKDGKSHFKWSPYLECENGSKPGWLVTLSAIYGLQP 294
DB 377 HEEG---GNRLNSSTAPRPLCTGDNINSVETPYMDYEHRLRISYNYLAVYSIAHALQD 433
QY 295 --NLVPERGVKVDINLQKVDIDQCSSDGSFSGTHKHLNNS-----ECMPIKG 342
DB 434 IYTCPLPG-RGLF---TNGSCADIKV--EAMQVLKHLRLHNFNTNMGEQVTFDECGDLVG 487
QY 343 LGFVL-----GAYECICKAG---FYHPGVLPVNNRRRGGPDQHS-----SPCFVQEDKYL 417
DB 488 NYSIINHLSPEDGSIVFKEVGYNYVAKKGERLFINEEKILWGSFREVFPNSCRDQC 547
QY 380 -GSTKDVSEE---AYVCLPCREG-----CPFCADD-----SPCFVQEDKYL 417
DB 548 AGTRKGIIEGPTCCFCEVCEPDGYSGETDASACDKCPDFFWSNENHTSCIAKEIEFLA 607
QY 418 -----LAIISFQGLCMMLDFVSMVYHFRKAKSIRASGLILLETILFGSILLYFPVVI 471
DB 608 WTEPFGLTALFAVLGIFLTAFLVGVFIKFRNTPIVKAATNRELSYLLFLSLCCFSSSLF 667

FILING DATE: 03-OCT-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/484,565
FILING DATE: 7-June-1995
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1079 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-986-8

Query Match 4.2%; Score 133.5; DB 2; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.00041;
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;
QY 235 HRRGNQGRGLGHSWRKDGKSHFKWSPPYLEGENGSKPGWLVTLSSAIYGLQP 294
DB 377 HEEG--GNRLNSSSTARPLCTGDNINSVETPYMDYEHRLRISYNYVLAVYSTAHALQD 433
295 --NLVPEPRGVKMDINLQKVDIDQCSSDGFSGTHKCHLNN-----ECMPIKG 342
DB 434 IYTCPLPG-RGLF--TNGSCADIKKV--EAMQVLKHLRLNFTNNMGQVTFDECGDLVG 487
QY 343 LGFVL-----GAYECICKAG--FYHPGVLPVNNFRRRGPDQHS----- 379
DB 488 NYSIINHLSPEDGSIVKVEGVYNNVAKGERLFINFEKTLWSGFSREVFPFNSCRDQC 547
QY 380 -GSTKDVSEE---AYVCLPREG-----CPFCADD-----SPCFVQEDKYL 417
DB 548 AGTRKGIIEGTEPTCCFCEVCEPDGEYSGETDASACDKCPDDFWNSHNHTSCIAKEIFELA 607
QY 418 -----LAITSFOGLMILLDFVMSLVYVHFRKAKSIRASGILILLETILFGSLLYFPVVI 471
DB 608 WTEPFGIALTLFAVLGIFLTAFLVGLFKFRNTPTVKTATNRELSVLLFLSLLCCFSSSLF 667
QY 472 LYEPSTPRCTLLRWRLIGFATVYGTVTKLHRVLKVLRSRTAQRIFYWTGGRVMRLA 531
DB 668 FIGEPQDWTCLRPQAGISVLCISILVKTNRVLLVFEAKIPTSFHRKWMGLNLQFL 727
QY 532 V-----ILVVFWEFLIGWTGSCVQNKLEQISLIGQKTSIDHLIFNMLIDRWDMYTAV 584
DB 728 VFLCTFMQILICIIWLVTAPSSY-RNHELE-----DEIIFITC---HEGSLMAL 773
QY 585 GMSVLSVY 592

DB 774 G--SLIGY 779
RESULT 5
US-08-353-784-8
; Sequence 8, Application US/08353784
; Patent No. 6011068
; GENERAL INFORMATION:
; APPLICANT: Edward F. Nemeth, Edward M.
; APPLICANT: Brown, Steven C. Hebert, Manuel
; APPLICANT: Bradford C. Van Wagenen, Manuel
; APPLICANT: F. Balandrin, Forrest H. Fuller,
; APPLICANT: Eric G. Delmar, and Scott T. Moe
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; MOLECULES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTSEQ
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,784
; FILING DATE: 9 December, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 8
; APPLICATION NUMBER: PCT/US/94/12117
; FILING DATE: 21 October, 1994
; APPLICATION NUMBER: U.S. 08/292,827
; FILING DATE: 23 August, 1994
; APPLICATION NUMBER: U.S. 08/141,248
; FILING DATE: 22 October, 1993
; APPLICATION NUMBER: U.S. 08/009,389
; FILING DATE: 23 February, 1993
; APPLICATION NUMBER: U.S. 08/017,127
; FILING DATE: 12 February, 1993
; APPLICATION NUMBER: U.S. 07/934,161
; FILING DATE: 21 August, 1992
; APPLICATION NUMBER: U.S. 07/834,044
; FILING DATE: 11 February, 1992
; APPLICATION NUMBER: U.S. 07/749,451
; FILING DATE: 23 August, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Heber, Sheldon O.
; REGISTRATION NUMBER: 38,179
; REFERENCE/DOCKET NUMBER: 209/069
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1079 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-353-784-8

Query Match 4.2%; Score 133.5; DB 3; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.00041;

Matches	93,	Conservative	58,	Mismatches	182,	Indels	95,	Gaps	10,
Qy	235	HRPGNPGRLGHSHRRKDLGGDKSHFKWSPPYLECEGSIYKPGWLVTLSIAIYGLQP	294						
Db	377	HEEG---GNRLNNSSTAPRLCTGDENINSVETPYMDYEHRLSRISVNYVLAVYSIAHALQD	433						
Qy	295	--NLVPEFRGVKVDIINLKQVDDOCSSDGSFGSTHCKHLANS-----ECMPIKG	342						
Db	434	IYTLCP--RGLF---TNGSCADIKKV--EAWQVLKHLRLHNTNMGGEQVTFDECGDLVG	487						
Qy	343	LGFLV-----GAYECICKAG---FYHFGVLPVNNFRRRGPQDHIS----	379						
Db	488	NYSTIINWHLSPEDGSIVFKEYGYNYVAKKGERLFINEEKILWSGFSREVFPFSCNRDQC	547						
Qy	380	--GSKVDYSEE---AYVCLPCREG-----CPTCADD-----SPCFVQEDKYL	417						
Db	548	AGTRKGIIEGPTCCFCECPDGESETDASADKCPDWSNENHTSCIAREIEFLA	607						
Qy	418	-----LAIISFOGLMLLDVSMVLVYVFRKAKSIRASGLILLETILFGSLILLYFPVVI	471						
Qy	608	WTEPFIALTFLAVLIGFLTAFLVGVFIKFRNTPIVKATNRELSYLLFLSLCCFSSSLF	667						
Qy	472	LYPSPSFRCTILLRWALLGGATVIGVITLKLHRYLVKVFLSRTQRIQRYMPTGGRRVMRLA	531						
Db	668	FIGEPQDMTCRLQPAFGISFVLCISILVKTNRVLLVFEAKIPTSFHRKWMGLNLQFL	727						
Qy	532	V-----ILLVWFELIGMTSSVQCNLEKOISLIGQKTSDDLHFLMCLDRWDYMTAV	584						
Db	728	VFLCTFMQILICILWLTAPSSY--RNHELE-----DEIFITC---HEGSLMAL	773						
Qy	585	GMWSLVSY	592						
Db	774	G--SLIGY	779						

RESULT 6
 US-08-484-719B-8
 ; Sequence 8, Application US/08484719B
 ; Patent No. 6031003
 ; GENERAL INFORMATION:
 ; APPLICANT: Edward F. Nemeth, Edward M.
 ; APPLICANT: Brown, Steven C. Hebert,
 ; APPLICANT: Bradford C. Van Wageningen,
 ; APPLICANT: Manuel F. Balandrin,
 ; APPLICANT: Forrest H. Fuller, Eric G.
 ; APPLICANT: Delmar, Scott T. Moe
 ; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
 ; MOLECULES
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Lyon & Lyon
 ; STREET: First Interstate World Center
 ; STREET: Suite 4700
 ; STREET: 633 West Fifth Street
 ; CITY: Los Angeles
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 90071
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: MS Word
 ; SOFTWARE: FASTSEQ for Windows Version 3.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/484,719B
 ; FILING DATE: 7 June, 1995
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/353,784
 ; FILING DATE: 9 December, 1994
 ; APPLICATION NUMBER: PCT/US/94/12117
 ; FILING DATE: 21 October, 1994
 ; APPLICATION NUMBER: U.S. 08/292,827

FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Douglas C. Murdock
REGISTRATION NUMBER: 37,549
REFERENCE/DOCKET NUMBER: 213/007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1079 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-719B-8

Query Match 4.2%; Score 133.5; DB 3; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.00041;
Matches 93; Conservative 58; Mismatches 182; Indels 95;

Qy	235	HRRPNOPGRGLGSHWRRKDGGLGDKSHFKWSPPYLECENGSYKKPGWLVTLSAIYGLQP	294
Db	377	HEEG---GNRLNNSAFRPLCTGDENINSVETPYMDYEHRLISYVNYLAVYSIAHALQD	433
Qy	295	--NLVPEFRGMVKVDINLQKVDIDQSSDGFSGTHKCHLNN-----ECMPIKG	342
Db	434	IYTCLPG-RGLF---TNGSCADIKV--EAWQVLKHLRHLFTNNMGEOVTFDECGLVG	487
Qy	343	LGFVL-----GAYECICKAG---FYHPGVLPVNNFRRRCPODHIS----	379
Db	488	NYSIINHLSPEDGISVFEKGVINYVAKKGERLFINEEKILWSGFSREVFPNSCRDQO	547
Qy	380	-GSTKDVSEE---AYVCLPCREG-----CPFCADD-----SPCFVQEDKYLR	417
Db	548	AGTRKGIIEGPTCCFCEVCEPDGEYSGETDASACDKCPDDFSWENHTSCIAKETFLA	607
Qy	418	-----LAISFQGLMLDFVMSLVVYHYHRKAKSTRASGLILLEFILGSLLLYPFVVI	471
Db	608	WTEFFGIALTLFAVLGILFATVLGVGFKEKRNPTPIKATNRELUSYLLLSLCCFSSSLF	667
Qy	472	LYFEPSTFRICLLRWARLLGFATVIGTVTLKHLRVLKVFLSRTAQRPTMYMTGGRVMRLA	531
Db	568	FIGEPQDWTCLRQPAFGISFVLICISILVKTNRVLLVPEAKIPTSFHRKMWGLNQFLL	727
Qy	532	V-----ILLVFWFLIGWTSVVCNLEKQISLIGOGKTSDDLHFNWCLMDRWDMYMTAV	584
Db	728	VFLCTFWOLLCIIWLVTAPSSY-RNHEUE-----DEIIIFTC---HEGSLMAL	773
Qy	585	GMWSLVSY	592
Db	774	G--SLIGY	779
Qy	592	SLIGY	779

RESULT 7
US-08-485-588-7
; Sequence 7, Application US/08485588
; Patent No. 5688938
; GENERAL INFORMATION:

APPLICANT: Edward M. Brown
APPLICANT: Steven C. Hebert
APPLICANT: Forrest H. Fuller
APPLICANT: James E. Garrett, Jr.
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,588
FILING DATE: 7 June, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 9
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/005
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-588-7

Query Match 4.0%; Score 127; DB 1; Length 1078;
Best Local Similarity 22.5%; Pred. No. 0.0018;
Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;
QY 311 QKVIDQSSDGFSGTHKC---HUNNSE-CMPKGLGVLGAYECICKAG---FYHPGV 363
DB 475 EQVTFDEC---GDLVGNYSIINWILSPEDGSIVFKEVGY---YNYAKKGERLFINEEK 527
QY 364 LPVNNFRFRGPDQH-----ISGSKVDVSEE-----AYVCLPCREG-----CPFCAD 404

Db 528 ILWSGFSREVPFNSCRDLAAGTRKGIIEGPTCFECVECPDGEYSDETDASACNKC PD 587
QY 405 D-----SPCFVQEDKYLR-----LAIISFOGLCMLLDFVSMVWYHFRKAKSIRASG 451
Db 588 DWSNENHTSCIAKEIEFLSWTEPGIALTFPAVLGIFLTAFLVGLGVFKFRWTPIVKATN 647
QY 452 LILLETLFGLSLLYFPVVILYFEPSTFRCLLLRWARLLGFATVYGTVTLKLRHVLKVF 511
Db 648 RELSYLLLSLLCCPSSSLFFIGEPQDWTCLRQPAFGISFVLCISCLVKTNRVLLVFE 707
QY 512 SRTAQRIPYMTGGRVNRMLAV-----ILLVVFELGWTSSV-CQMLEKQISLI----- 559
Db 708 AKIPTSFHRKMWMLQLQFLVFLCTFMQIVTCVIMLVYTAPSSYRNQDELEDEIIFITCHE 767
QY 560 GQKTSDDLIFNMCLI 575
Db 768 GSLMALGFLIGYTCLL 783
RESULT 8
US-08-484-565-7
Sequence 7, Application US/08484565
Patent No. 5763569
GENERAL INFORMATION:
APPLICANT: Edward M. Brown
APPLICANT: Steven C. Hebert
APPLICANT: James E. Garrett, Jr.
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,565
FILING DATE: 7 June, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 9
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/006

TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-565-7

Query Match 4.0%; Score 127; DB 1; Length 1078;
Best Local Similarity 22.5%; Pred. No. 0.0018;
Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;

QY 311 QKVDIDOCSSDGFSGTHKC---HLNNS-CMPKIGLGVLGAYECICKAG---FYHPGV 363
475 EQVTFDEC---GDLVGNYSIINHLSPEDSIVFKEVGY---YNYAKKGERLFINEEK 527
364 LPVNNFRRRGPDOH-----ISGSTKDYSEE---AYVCLPCREG-----CPFCAD 404
528 ILWSGFSREVPFNSCRDCLAGTRKGIIEGPTCCFCEVCEPDGEYSDETDASACNCPD 587
QY 405 D-----SPCFVQEDKYLK-----LAIISFOGLCMLLDFVSMVYVYHFRKAKSIRASG 451
588 DFWSNENHTSCIAKEIEFLSWTEPFGLIALTLFAVLGIFLTAFLGVFKFRNTPIVKATN 647
QY 452 LILLETILFGLSLLYFPVILYFPEPSTFCILLRWALLGFAIVYGVVTKLHRVLKVL 511
648 RELSYLLLSLLCCFSSSLFFIGEPQDWTCLRPAPAGISFVLCISCLVKTNRVLLVFE 707
QY 512 SRTAQRIPTYMTGGVRMRLAV-----ILLVFWFLIGHTSSV-CQMLEKQISLI----- 559
708 AKIPTSFRKRWGMLNLOFLVFLCTFMQIVICVILWLTAPPSSYRNQOELEDEIIFITCHE 767
QY 560 GQKTSDDLIFNMCLI 575
768 GSLMALGFLIGYTCLL 783

RESULT 9
US-08-480-751-7
Sequence 7, Application US/08480751
Patent No. 5858684
GENERAL INFORMATION:
APPLICANT: Edward F. Nemeth
APPLICANT: Edward M. Brown
APPLICANT: Steven C. Hebert
APPLICANT: Forrest H. Fuller
APPLICANT: James E. Garrett, Jr.
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,751
FILING DATE: 7 June, 1995
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 9
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-751-7

Query Match 4.0%; Score 127; DB 2; Length 1078;
Best Local Similarity 22.5%; Pred. No. 0.0018;
Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;

QY 311 QKVDIDOCSSDGFSGTHKC---HLNNS-CMPKIGLGVLGAYECICKAG---FYHPGV 363
475 EQVTFDEC---GDLVGNYSIINHLSPEDSIVFKEVGY---YNYAKKGERLFINEEK 527
364 LPVNNFRRRGPDOH-----ISGSTKDYSEE---AYVCLPCREG-----CPFCAD 404
528 ILWSGFSREVPFNSCRDCLAGTRKGIIEGPTCCFCEVCEPDGEYSDETDASACNCPD 587
QY 405 D-----SPCFVQEDKYLK-----LAIISFOGLCMLLDFVSMVYVYHFRKAKSIRASG 451
588 DFWSNENHTSCIAKEIEFLSWTEPFGLIALTLFAVLGIFLTAFLGVFKFRNTPIVKATN 647
QY 452 LILLETILFGLSLLYFPVILYFPEPSTFCILLRWALLGFAIVYGVVTKLHRVLKVL 511
648 RELSYLLLSLLCCFSSSLFFIGEPQDWTCLRPAPAGISFVLCISCLVKTNRVLLVFE 707
QY 512 SRTAQRIPTYMTGGVRMRLAV-----ILLVFWFLIGHTSSV-CQMLEKQISLI----- 559
708 AKIPTSFRKRWGMLNLOFLVFLCTFMQIVICVILWLTAPPSSYRNQOELEDEIIFITCHE 767
QY 560 GQKTSDDLIFNMCLI 575
768 GSLMALGFLIGYTCLL 783
RESULT 10
US-08-943-986-7
Sequence 7, Application US/08943986
Patent No. 5962314
GENERAL INFORMATION:
APPLICANT: Edward M. Brown

APPLICANT: Steven C. Hebert
APPLICANT: James E. Garrett, Jr.
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,986
FILING DATE: 03-OCT-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/484,565
FILING DATE: 7-June-1995
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/006
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-986-7

Query Match 4.0%; Score 127; DB 2; Length 1078;
Best Local Similarity 22.5%; Pred. No. 0.0018;
Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;
Qy 311 OKVIDOCSSDGSFSTHCK---HLNNE-CMPKGLGFLVGLAYECICKAG---FYHFGV 363
Db 475 EQVTDEC---GDLVGNYSIINHLSPDGSIVFKEVGY-----YNVYAKGRERUFINEEK 527
Qy 364 LPVNNFRRRGPDQH-----ISGSTRKDVSEE-----AYVCLPCREG-----CPFFCAD 404
Db 528 ILWSGFSREVPFNSRDLAGTRKGIIEGPTCCFECVCPDGEYSDETASACNKCPCD 587

Qy 405 D-----SPOFVQEDKYLR-----LAIISQGLCMLLDFFVSMVLVYVHFRKAKSIRASG 451
Db 588 DFWSNENHSTCIAKEIEFLSWTEPEFGIALTLFAVLGIELTFAVLGVFIKFRNTPIVKATN 647
Qy 452 LILLETLFGSLLYFPVVLVYFEPSTFCILLRWALLGFATVYGTVLKLRHVLKVF 511
Db 648 RELSYLLFLSLCCFSSSLFFIGEPQDWTCLRQPAFGISFVLCISILVKTNRVLLVFE 707
Qy 512 SRTAQRIPYMTGGVRMLAV-----ILLVWFELGTWSSV-CQNEKQISLI----- 559
Db 708 AKIPTSFHRKWWGLNLQFLVFLCTFMQIVICVIMLYTAPSSYRNQOELEDEIIFITCHE 767
Qy 560 GOGKTSDDLIFNMCLI 575
Db 768 GSLMALGFLGYTCLL 783
RESULT 11
US-08-353-784-7
Sequence 7, Application US/08353784
Patent No. 6011068
GENERAL INFORMATION:
APPLICANT: Edward F. Nemeth, Edward M.
APPLICANT: Brown, Steven C. Hebert,
APPLICANT: Bradford C. Van Wagenen, Manuel
APPLICANT: F. Balandrin, Forrest H. Fuller,
APPLICANT: Eric G. DelMar, and Scott T. Moe
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,784
FILING DATE: 9 December, 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 8
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 209/069
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600

TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-353-784-7

Query Match 4.0%; Score 127; DB 3; Length 1078;
Best Local Similarity 22.5%; Pred. No. 0.0018;
Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;

QY 311 QKVIDOCSSDGFSGTHKC---HLNNS-CMPIKGLGVLGAYECICKAG---FYHPGV 363
Db 475 EQVTFDEC---GDLVGNYSIINWHLSPEDGSIVFKEVGY---YNYVAKKGERLFINEEK 527
364 LPVNNFRRRGPDOH-----ISGSKTKDVSEE---AYVCLPCREG-----CPFCAD 404
528 ILWSGFSREVFPFNSCRDCLAGTRKGIIEGPTCCFCEVCPDGEYSDETDASACNKKCPD 587
QY 405 D-----SPCFVQEDKYLRL-----LAIISFOGLCMLDFVSMVYHFRKAKSIASG 451
Db 588 DFWSNENHTSCIAKEIEFLSWTEPFGIALTLFAVLGIFLTAFLVGVFKFRNTPIVKATN 647
QY 452 LILLETILFGLSLLYPVVTLYPEPSTFRICILLRWARLLGFAIVGTIVTLKLRVLYKVL 511
Db 648 RELSYLLFLSLCCFSSSLFFIGEPQDWTCLRPQAFGISFVLCISILYKTRNVLVFE 707
QY 512 SRTAQRIPTYMTGVRMMLAV-----ILLVFWFLIGWTSV-CONLEKQISLI----- 559
Db 708 AKIPTSFRKRWGGLNQFLVFLCTFMQIVICVILWLYTAPPSSVRNOELEDIEIFITCHE 767
QY 560 GQKTSDDLIFNMCLI 575
Db 768 GSLMALGFLIGYTCLL 783

RESULT 12
US-08-484-719B-7
Sequence 7, Application US/08484719B
Patent No. 6031003
GENERAL INFORMATION:
APPLICANT: Edward F. Nemeth, Edward M.
APPLICANT: Brown, Steven C. Hebert,
APPLICANT: Bradford C. Van Wagenen,
APPLICANT: Manuel F. Balandrin,
APPLICANT: Forrest H. Fuller, Eric G.
APPLICANT: Delmar, Scott T. Moe
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS Word
SOFTWARE: FASTSEQ for Windows Version 3.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,719B
FILING DATE: 7 June, 1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Douglas C. Murdock
REGISTRATION NUMBER: 37,549
REFERENCE/DOCKET NUMBER: 213/007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-719B-7

Query Match 4.0%; Score 127; DB 3; Length 1078;
Best Local Similarity 22.5%; Pred. No. 0.0018;
Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;

QY 311 QKVIDOCSSDGFSGTHKC---HLNNS-CMPIKGLGVLGAYECICKAG---FYHPGV 363
Db 475 EQVTFDEC---GDLVGNYSIINWHLSPEDGSIVFKEVGY---YNYVAKKGERLFINEEK 527
364 LPVNNFRRRGPDOH-----ISGSKTKDVSEE---AYVCLPCREG-----CPFCAD 404
528 ILWSGFSREVFPFNSCRDCLAGTRKGIIEGPTCCFCEVCPDGEYSDETDASACNKKCPD 587
QY 405 D-----SPCFVQEDKYLRL-----LAIISFOGLCMLDFVSMVYHFRKAKSIASG 451
Db 588 DFWSNENHTSCIAKEIEFLSWTEPFGIALTLFAVLGIFLTAFLVGVFKFRNTPIVKATN 647
QY 452 LILLETILFGLSLLYPVVTLYPEPSTFRICILLRWARLLGFAIVGTIVTLKLRVLYKVL 511
Db 648 RELSYLLFLSLCCFSSSLFFIGEPQDWTCLRPQAFGISFVLCISILYKTRNVLVFE 707
QY 512 SRTAQRIPTYMTGVRMMLAV-----ILLVFWFLIGWTSV-CONLEKQISLI----- 559
Db 708 AKIPTSFRKRWGGLNQFLVFLCTFMQIVICVILWLYTAPPSSVRNOELEDIEIFITCHE 767
QY 560 GQKTSDDLIFNMCLI 575
Db 768 GSLMALGFLIGYTCLL 783

RESULT 13
US-08-485-588-5
Sequence 5, Application US/08485588
Patent No. 5688938
GENERAL INFORMATION:
APPLICANT: Edward M. Brown
APPLICANT: Steven C. Hebert
APPLICANT: Forrest H. Fuller

OM of: US-09-775-181-4 to: GenEmbl.* out_format : pfs

Date: Nov 3, 2001 1:32 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame*_p2n.model -DEV=xlh
-O/cgnt2/_USPTO.spool/US09775181/runat_30102001_161056_13810/app_query.fasta_1.1955
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOPC=0.000 -LOPEXT=0.500 -OGAPOP=4.500
-OGAEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09775181@cgnl_1.11947 -NCPU=6
-ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY
-WAIT -THREADS=1

arch information block:

Query: US-09-775-181-4
Query length: 599
Database: GenEmbl.*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 4100.330000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_hlg21:AL358932	+	1607.00	1880.63	1.6e-96	11117	! AL358932 Homo sapiens chrom
gb_hlg20:AL139821	+	1607.00	1875.72	3.1e-96	199526	! AL139821 Homo sapiens chrom
gb_hlg20:AL355587	+	489.00	540.44	7.3e-22	151218	! AL355587 Homo sapiens chrom
gb_hlg21:AL359974	-	409.00	445.14	1.5e-16	143878	! AL359974 Homo sapiens chrom
gb_hlg20:AL161654	-	403.00	437.09	4.4e-16	159996	! AL161654 Homo sapiens chrom
gb_hlg5:AC017271	-	259.50	287.92	8.4e-08	10892	! AC017271 Drosophila melanoga
gb_hlg1:AC009845	-	259.50	264.22	1.8e-06	183766	! AC009845 Drosophila melanoga
gb_hlg1:AC009357	+	259.50	259.39	3.3e-06	327209	! AC009357 Drosophila melanoga
gb_hlg20:AL354976	+	202.00	195.28	0.0122	186964	! AL354976 Homo sapiens chrom
gb_hlg22:AL445244	-	200.00	194.18	0.0140	160390	! AL445244 Homo sapiens chrom
gb_hlg4:AC014084	-	183.50	174.99	0.1646	150133	! AC014084 Drosophila melanoga
gb_hlg1:AC007772	-	183.50	173.17	0.2079	186607	! AC007772 Drosophila melanoga
gb_hlg2:AE003732	-	183.50	172.71	0.2205	197071	! AC007772 Drosophila melanoga
gb_hlg2:AE003732	-	183.50	171.26	0.2657	234369	! AE003732 Drosophila melanoga
gb_hlg3:DMX99675	+	164.00	183.61	0.0845	3325	! X99675 D.melanogaster mRNA fol
gb_hlg9:HSPCAR1	+	161.00	180.25	0.0838	3234	! X81086 H. sapiens Pcar1 gene.
gb_hlg10:HSPCAR1	+	161.00	180.25	0.0838	3234	! S83176 Homo sapiens calcium-se
gb_hlg10:HSCASR	+	161.00	179.93	0.0873	3361	! D50855 Human mRNA for Ca-sensi
gb_hlg10:HSU20759	+	161.00	178.94	0.0992	3783	! U20759 Human parathyroid cell
gb_hlg1:AR012624	+	161.00	178.88	0.0999	3809	! AR012624 Sequence 3 from paten
gb_hlg1:AR028467	+	161.00	178.88	0.0999	3809	! AR028467 Sequence 3 from paten
gb_hlg1:AR078219	+	161.00	178.88	0.0999	3809	! AR078219 Sequence 3 from paten
gb_hlg2:175053	+	161.00	178.88	0.0999	3809	! I75053 Sequence 3 from paten
gb_hlg4:AC014500	+	160.00	163.45	0.7228	20791	! AC014500 Drosophila melanoga
gb_hlg1:AC009257	+	160.00	143.72	9.07	218565	! AC009257 Drosophila melanoga
gb_hlg1:AE003461	+	160.00	141.11	12.69	298616	! AE003461 Drosophila melanoga
gb_hlg1:AE003461	+	152.50	166.02	0.5199	5249	! S67307 Ca(2+)-sensing receptor
gb_hlg1:AR012622	+	152.50	165.98	0.5227	5275	! AR012622 Sequence 1 from paten
gb_hlg1:AR028465	+	152.50	165.98	0.5227	5275	! AR028465 Sequence 1 from paten
gb_hlg1:AR078217	+	152.50	165.98	0.5227	5275	! AR078217 Sequence 1 from paten
gb_hlg2:175051	+	152.50	165.98	0.5227	5275	! I75051 Sequence 1 from paten
gb_hlg1:AR012623	+	150.00	163.43	0.7251	5006	! AR012623 Sequence 2 from paten
gb_hlg1:AR028466	+	150.00	163.43	0.7251	5006	! AR028466 Sequence 2 from paten
gb_hlg1:AR078218	+	150.00	163.43	0.7251	5006	! AR078218 Sequence 2 from paten
gb_hlg2:175052	+	150.00	163.43	0.7251	5006	! AR078218 Sequence 2 from paten
gb_hlg2:G28586	+	150.00	163.42	0.7255	5009	! I75052 Sequence 2 from paten
gb_hlg10:HSU20760	+	150.00	163.42	0.7255	5009	! G28586 Human STS SHCC-35350, s
gb_hlg1:AR084671	+	148.50	163.51	0.7171	4000	! U20760 Human extracellular cal
gb_hlg4:AC014621	-	148.00	150.23	3.94	18160	! AR084671 Sequence 2 from paten
gb_hlg2:AE003846	-	148.00	126.75	79.99	298469	! AC014621 Drosophila melanoga

gb_roi:AF110178 + 145.50 158.84 1.31 4550 ! AF110178 Mus musculus calci
gb_roi:AF128842 - 144.50 159.68 1.17 3569 ! AF128842 Mus musculus extra
gb_in3:LMFL6520 - 144.50 139.33 15.95 40438 ! AL446005 Leishmania major
gb_ov:AB008857 + 144.00 159.00 1.28 3606 ! AB008857 Fugu rubripes gene
gb_in2:AF318274 + 144.00 158.30 1.40 3918 ! AF318274 Drosophila melanog

seq_name: gb_hlg21:AL358932

seq_documentation_block:

LOCUS AL358932 111117 bp DNA HTG 23-JAN-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-395P8, *** SEQUENCING IN
PROGRESS ***, 25 unordered pieces.

ACCESSION AL358932 GI:9930957

VERSION HTG; HTGS_PHASE1.

KEYWORDS human.

SOURCE

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Sims,S.

TITLE Direct Submission

JOURNAL

COMMENT

Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 27, 2000 this sequence version replaced gi:9926667.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA395P8

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 90261 bases at least Q40

Consensus quality: 92226 bases at least Q30

Consensus quality: 104490 bases at least Q20

Insert size: 108717; sum-of-contigs

Insert size: 178772; 24.0% error; agarose-fp

Quality coverage: 1.84x in Q20 bases; sum-of-contigs Quality

coverage: 1.42x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 25 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 2287: contig of 2287 bp in length

* 2288 2387: gap of 100 bp

* 2388 6317: contig of 3930 bp in length

* 6318 6417: gap of 100 bp

* 6418 9725: contig of 3308 bp in length

* 9726 9825: gap of 100 bp

* 9826 16607: contig of 6782 bp in length

* 16608 16707: gap of 100 bp

* 16708 20159: contig of 3452 bp in length

* 20160 20259: gap of 100 bp

* 20260 23641: contig of 3382 bp in length

* 23642 23741: gap of 100 bp

* 23742 27399: contig of 3658 bp in length

* 27400 27499: gap of 100 bp

* 27500 32754: contig of 5255 bp in length

* 32755 32854: gap of 100 bp

* 32855 35240: contig of 2386 bp in length

* 35241 35340: gap of 100 bp

* 35341 40948: contig of 5608 bp in length

* 40949 41048: gap of 100 bp


```
83853 CAGGACGACCTGGATGGTACCAGCGCTGGTGTGGAGCCTTCTGGAGGG 83902
167 yGluProSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeus 184
83903 CGAGCCACGACTCTCCGGGGCGGCATCACCTTCAGCACCGATTGCTGT 83952
184 erAlaProAlaProGlnValPheLeuGlnAlaThrArgGluGluSerArg 200
83953 CGGCACCGGGCCACAGGTCTTCCTCCAGGCGCACGCGGAGGAGCGCG 84002
201 ileLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaTh 217
84003 ATCCTGCTCCAGACCTGCTCTCTCCGACCCCACTGGCCAGGCCAC 84052
217 rLeuGluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisL 234
84053 TCTGGAGACCGAGTGTCTCCACGGCTCTCCGGCGCAAGTGGAGGCCACT 84102
234 euHisArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrp 250
84103 TACACCGCGCGCGGCCCAATACAGGGGCCCCGGGGCTGGGCCACAGCTGG 84152
251 ArgArgLysAspGlyLeuGlyAspLysSerHisPheLysTrpSerPr 267
84153 CGGCGCAAGGAGCGGCTCGCGGGGACAGAGCCACTTCAAGTGTCTCC 84202
267 oProTyrLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeuValT 284
84203 GCCTTATCTGGAGTGCAGAGACGGGAGTTACAAGCCCGGTGGTGGTTA 84252
284 hrLeuSerSerAlaIleThrGlyLeuGlnProAsnLeuValProGluPhe 300
84253 CTCCTTCTCTGCCATCTACGGGTTCCAGCCCTAACCTGGTCCCGGAATC 84302
301 Arg 301
111
84303 AGG 84305
seq_name: gb_btg20:AL139821
seq_documentation_block:
LOCUS AL139821 199526 bp DNA HTG 17-MAR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-80K21, *** SEQUENCING IN
PROGRESS ***, 6 unordered pieces.
ACCESSION AL139821
VERSION AL139821.7 GI:13396382
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Plumb,B.
Direct Submission
Submitted (15-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 20, 2001 this sequence version replaced gi:10185474.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA80K21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 197300 bases at least Q40
Consensus quality: 198299 bases at least Q30
Consensus quality: 198676 bases at least Q20
Insert size: 199026; sum-of-contigs
```

```
Insert size: 144047; 11.6% error; agarose-fp
Quality coverage: 7.42x in Q20 bases; sum-of-contigs Quality
coverage: 10.32x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 37012: contig of 37012 bp in length
* 37013 37112: gap of 100 bp
* 37113 163560: contig of 126448 bp in length
* 163561 163660: gap of 100 bp
* 163661 171563: contig of 7903 bp in length
* 171564 171663: gap of 100 bp
* 171664 186786: contig of 15123 bp in length
* 186787 186886: gap of 100 bp
* 186887 196057: contig of 9171 bp in length
* 196058 196157: gap of 100 bp
* 196158 199526: contig of 3369 bp in length.
FEATURES
            Location/Qualifiers
            source
            1..199526
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="10"
            /clone="RP11-80K21"
            /clone_lib="RPC1-11.1"
            1..37012
            /note="assembly_fragment:00668"
            fragment_chain:1
            clone_end:T7
            vector_side:left"
            37113..163560
            /note="assembly_fragment:02214"
            fragment_chain:1
            clone_end:SP6
            vector_side:right"
            163661..171563
            /note="assembly_fragment:00510"
            171664..186786
            /note="assembly_fragment:01538"
            186887..196057
            /note="assembly_fragment:01931"
            196158..199526
            /note="assembly_fragment:03255"
BASE COUNT 57428 a 38898 c 39349 g 63347 t 504 others
ORIGIN
alignment_scores:
            Quality: 1607.00      Length: 301
            Ratio: 5.339          Gaps: 0
            Percent Similarity: 100.000      Percent identity: 100.000
alignment_block:
US-09-775-181-4 x AL139821
..
Align seg 1/1 to: AL139821 from: 1 to: 199526
1 MetGlyAlaMetAlaTyrProLeuLeuLeuCysLeuLeuAlaGlnle 17
|||||
97158 ATGGAGGCGCATGGCTTACCCCTTACTCTCTGCTTCTGCTCAGCT 97207
17 uclyLeuGlyAlaValGlyAlaSerArgAspProGlnGlyArgProAsps 34
|||||
97208 GGGATTGGAGCTGTTGGCCAGCGCGCACCCCAAGGACGCGCGGATT 97257
34 erProArgGluArgThrProLysGlyLysProHisAlaGlnGlnProGly 50
|||||
97258 CCCCTCGAGAGAGACCCCGAAGGGGAAGCGCACGCGCCAGCGCGGT 97307
```

```
51 ArgAlaSerAlaSerAspSerSerAlaProTrpSerArgSerThrAspG1 67
|||||
97308 CGAGCCCTCTCGCTCGGACTCCTCGGCTCCCTGGAGCGCTCCACCGATGG 97357
|||||
67 yThrIleLeuAlaGlnLysLeuAlaGluGluValProMetAspValAlaS 84
|||||
97358 CACCATCTTGGCGAGAAATCGCCGAGGAGTGCCTCATGACGTGGCCT 97407
|||||
84 eTyrLeuTyrThrGlyAspSerHisGlnLeuLysArgAlaAsnCysSer 100
|||||
97408 CTTACCTCTACACGGGGACTCCACAGCTGAAGGAGGACCACTGCTCC 97457
|||||
101 GlyArgTyrGluLeuAlaGlyLeuProGlyLysTrpProAlaLeuAlaSe 117
|||||
97458 GGCGCGCTACAGTGGCGGCGCTCGCGGGAAGTGGCCAGCCCTGGCCAG 97507
|||||
117 rAlaHisProSerLeuHisArgAlaLeuAspThrLeuThrHisAlaThrA 134
|||||
97508 CGCGCACCCCTCTCTGACCGGCGCTGGACACACTGACACAGCCACCA 97557
|||||
134 snPheLeuAsnValMetLeuGlnSerAsnLysSerArgGluGlnAsnLeu 150
|||||
97558 ACTCTCTACAGTGTGCTGCAGAGCAATAGTTCGGGGAGCAGAACTTG 97607
|||||
151 GlnAspAspLeuAspTrpTyrGlnAlaLeuValTrpSerLeuLeuGluG1 167
|||||
97608 CAGGACGACCTGGATTGGTACCAGCGCTGCTGTGGAGCCTTCTGGAGGG 97657
|||||
167 yGluProSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeuS 184
|||||
97658 CGAGCCGAGCATCTCCGGCGGCGCATACCTTCAGCAGCGATTCGCTGT 97707
|||||
184 eAlaProAlaProGlnValPheLeuGlnAlaThrArgGluGluSerArg 200
|||||
97708 CCGCAGCGGCGCCACAGGTCTTCTCTCAGGCCACGCGCGGAGGAGCCGC 97757
|||||
201 IleLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaTh 217
|||||
97758 ATCTGTCTCCAGACCTGTCTCTCCGACCCACCTGCGCCAAAGCCAC 97807
|||||
217 rLeuGluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisL 234
|||||
97808 TCTGGAGACGAGTGGTTCACGGCTCCGGCCCAAGTGGAGGAGCCCACT 97857
|||||
234 euHisArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrp 250
|||||
97858 TACACCGCGCGGCCCCATCAGGGCGCCGGGCGCTGGGCCACAGCTGG 97907
|||||
251 ArgArgLysAspGlyLeuGlyAspLysSerHisPheLysTrpSerPr 267
|||||
97908 CGCGCAAGCAGCGGCTCGCGGGGACAGAGCCACTTCAAGTGGTCTCC 97957
|||||
267 oProTyrLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeuValT 284
|||||
97958 GCCTTATCTGGAGTGGAGACGGAGTTCACAGCCCGGGTGGCTGGTTA 98007
|||||
284 hrLeuSerSerAlaIleTyrGlyLeuGlnProAsnLeuValProGluPhe 300
|||||
98008 CTCCTTCTCTGCATCTACGGGTTCAGCCTAACCTGTGTCGGGATTC 98057
|||||
301 Arg 301
|||
98058 AGG 98060
```

seq_name: gb_hg20:AL355587

seq_documentation_block:

LOCUS AL355587 151218 bp DNA HTG 15-APR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-561H23, *** SEQUENCING IN
PROGRESS ***, 9 unordered pieces.
ACCESSION AL355587
VERSION AL355587.8 GI:13660940

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

HTG; HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 151218)
Direct Submission
Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:13624988.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA561H23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 147891 bases at least Q40
Consensus quality: 149178 bases at least Q30
Consensus quality: 149893 bases at least Q20
Insert size: 150418; sum-of-contigs
Insert size: 151193; 7.4% error; agarose-fp
Quality coverage: 4.69x in Q20 bases; sum-of-contigs Quality
coverage: 5.08x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 21608: contig of 21608 bp in length
* 21609 21708: gap of 100 bp
* 21709 48919: contig of 27211 bp in length
* 48920 49019: gap of 100 bp
* 49020 57539: contig of 8520 bp in length
* 57540 57639: gap of 100 bp
* 57640 69961: contig of 12322 bp in length
* 69962 70061: gap of 100 bp
* 70062 101299: contig of 31238 bp in length
* 101300 101399: gap of 100 bp
* 101400 114233: contig of 12834 bp in length
* 114234 114333: gap of 100 bp
* 114334 126584: contig of 12251 bp in length
* 126585 126684: gap of 100 bp
* 126685 138216: contig of 11532 bp in length
* 138217 138316: gap of 100 bp
* 138317 151218: contig of 12902 bp in length.

FEATURES
source

1. 151218
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-561H23"
/clone_lib="RPCI-11.2"
1. 21608
/note="assembly_fragment:00445
fragment_chain:1
clone_end:T7
vector_side:left
21709. 48919
/note="assembly_fragment:00191
fragment_chain:1
49020. 57539
/note="assembly_fragment:01755

misc_feature

misc_feature

misc_feature

66414	TTGAACCTTTTGGCTTTTCTAGGCTTTTGAAGGTGTTTCTTTTACGAAACGC	66463
515	aGlnAqilleProTyrMetThrGlyGlyArgValMetArgMetLeuAlav	532
66464	TCACAGNATTCATATATGACTTGGCGGACGGGTATGAGGATGCTGCAG	66513
532	alIleLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerSerVal	548
66514	TAATACTCTTGGTAGTGTGTTTGGTTTCTCATTTGGCTGGACTTCATCTGTG	66563
549	CysGlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSe	565
66564	TGGCAGTATTGGAGAAACAGATTTCACTTATTTGGCCATGGGAAACATC	66613
565	rAspHisLeuIlePheAsnMetCysLeuIleAspArgTirPaspTyrMetT	582
66614	CGATCACCCTCATCTTCAATATGTGCTCATTTGACCGTGGGACTACATGA	66663
582	hrIalValGlyMetTrpSerLeuValSerTyrAspGlyLeuThrIlePhe	598
66664	CAGCAGTTGGTAGTGTGCTCAGTGTTCGTATGATGCTTACCATTTT	66713
599	Gln 599	
66714	CAG 66716	
seq_name: gb_htg21:AL359974		
seq_documentation_block:		
LOCUS	AL359974 143878 bp DNA HTG	23-JAN-2001
DEFINITION	Homo sapiens chromosome 10 clone RP13-236A4, *** SEQUENCING IN PROGRESS ***, 25 unordered pieces.	
ACCESSION	AL359974	
VERSION	AL359974.4 GI:9864577	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 143878)	
TITLE	Sims, S.	
JOURNAL	Direct Submission	
	Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk	
	requests: clonerquest@sanger.ac.uk	
COMMENT	On Aug 22, 2000 this sequence version replaced gi:9795088.	
	----- Genome Center	
	Center: Sanger Centre	
	Center code: SC	
	Web site: http://www.sanger.ac.uk	
	Contact: humquery@sanger.ac.uk	
	----- Project Information	
	Center project name: BB236A4	
	----- Summary Statistics	
	Assembly program: XGAP4; version 4.5	
	Sequencing vector: plasmid; L08752; 100% of reads	
	Chemistry: Dye-terminator Big Dye; 100% of reads	
	Consensus quality: 130224 bases at least Q40	
	Consensus quality: 135875 bases at least Q30	
	Consensus quality: 138890 bases at least Q20	
	Insert size: 141478; sum-of-contigs	
	Insert size: 205930; 23.5% error; agarose-fp	
	Quality coverage: 3.17x in Q20 bases; sum-of-contigs Quality	
	coverage: 2.50x in Q20 bases; agarose-fp	

	* NOTE: This is a 'working draft' sequence. It currently consists of 25 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	


```

89183 ACAGCCCATGCTCGTCCAGGAAGATAAGTATTTAGCACTTGCCATCATC 89134
|||||
422 SerPheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeuValVa 438
|||||
89133 TCCTTCCAAGGCGTGTATGCTGCTGCGACTTCGTTAGCATGCTGCTGCT 89084
|||||
438 lTyRHlsPheArgLysAlaLysSerIleArgAlaSerGlyLeuLeuLeu 455
|||||
89083 CTACCACTTTCGCAAGCAAG...GTAACCCAGGAGCAACCTGGTT... 89041
|||||
455 euGluThrIleLeuPheGlySerLeuLeuLeuTyRPhProValValIle 471
|||||
89040 .....ATGATCCTGTATTACAGACCACTCT 89014
|||||
472 LeuTyRPhGluProSerThrPhe 479
|||||
89013 TATCTTTTAAAGCAACAGAAATT 88990
|||||
seq_name: gb_hgt20:AL161654
seq_documentation_block:
LOCUS AL161654 159596 bp DNA HTG 11-APR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-59G22, *** SEQUENCING IN
PROGRESS ***, 8 unordered pieces.
ACCESSION AL161654
VERSION AL161654.8 GI:13620309
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Burton, J.
Direct Submission
Submitted (10-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
On Apr 12, 2001 this sequence version replaced gi:13567947.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA59G22
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 157954 bases at least Q40
Consensus quality: 158428 bases at least Q30
Consensus quality: 158729 bases at least Q20
Insert size: 158896; sum-of-contigs
Insert size: 159333; 6.0% error; agarose-fp
Quality coverage: 6.69x in Q20 bases; sum-of-contigs Quality
coverage: 6.67x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 24276: contig of 24276 bp in length
* 24277 24376: gap of 100 bp
* 31732 31732: contig of 7356 bp in length
* 31733 31832: gap of 100 bp
* 31833 67342: contig of 3510 bp in length
* 67343 67442: gap of 100 bp
* 67443 70588: contig of 3146 bp in length

```

```

* 70589 70688: gap of 100 bp
* 70689 81295: contig of 10607 bp in length
* 81296 81395: gap of 100 bp
* 81396 111898: contig of 30503 bp in length
* 111899 111998: gap of 100 bp
* 111999 126181: contig of 14183 bp in length
* 126182 126281: gap of 100 bp
* 126282 159596: contig of 33315 bp in length.
FEATURES
    source
        1..159596
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="10"
            /clone="RP11-59G22"
            /clone_lib="RFC1-11.1"
            1..24276
                /note="assembly_fragment:01962"
                fragment_chain:1
                    clone_end:17
                    vector_side:left
                24377..31732
                    /note="assembly_fragment:01615"
                    fragment_chain:1
                31833..67342
                    /note="assembly_fragment:02112"
                    fragment_chain:1
                67443..70588
                    /note="assembly_fragment:00781"
                    fragment_chain:1
                70689..81295
                    /note="assembly_fragment:00892"
                    fragment_chain:1
                81396..111898
                    /note="assembly_fragment:02522"
                    fragment_chain:1
                111999..126181
                    /note="assembly_fragment:01840"
                    fragment_chain:1
                126282..159596
                    /note="assembly_fragment:01914"
                    fragment_chain:1
            clone_end:SP6
            vector_side:right
        54172 a 29996 c 28660 g 46068 t 700 others
    ORIGIN
        alignment_scores:
            Quality: 403.00      Length: 108
            Ratio: 4.478        Gaps: 2
            Percent Similarity: 83.333 Percent Identity: 72.222
        alignment_block:
            US-09-775-181-4 x AL161654/rev ..
            Align seg 1/1 to reverse of: AL161654 from: 1 to: 159596
            372 ArGgLyProAspGlnHisIleSerGlySerThrLysAspValSerGluG1 388
            102605 AGGGGTCGGCAGTATTTTCAGGAAGTACAAAGATGTGTCAGAGA 102556
            388 uALatYrValCysLeuProCysArgGluGlyCysProPheCysAlaAspA 405
            102555 AGCCTATGTCCTACCTTGCAGGAGGCGTCCCTCTCTGTGCTGATG 102506
            405 sPSerProCysPheValGlnGluAspLysTyrlleuArgIleuAlaIle 421
            102505 ACAGCCCATGCTCGTCCAGGAAGATAAGTATTTAGCACTTGCCATCATC 102456
            422 SerPheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeuValVa 438
            102455 TCCTTCCAAGGCGTGTATGCTGCTGCGACTTCGTTAGCATGCTGCTGCT 102406

```

```
438 lTyrHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuIleLeuL 455
|||||
102405 CTACCACTTTGCAAGCAAG...GTAAACCCAGACCTGGTT... 102363
455 euGluThrIleLeuPheGlySerLeuLeuLeuTyrPheProValValIle 471
|||||
102362 .....ATGATCCCGTATACAGACGACCTCT 102336
472 LeuTyrPheGluProSerThrPhe 479 HTG 09-DEC-1999
|||||
102335 TATCTTTTAAAGCAACAGATTT 102312
seq_name: gb_htg5:AC017271

seq_documentation_block:
LOCUS AC017271 10892 bp DNA HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces
ACCESSION AC017271
VERSION AC017271.1 GI:6553715
FEATURES
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 10892)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10210111 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Source
1. .10892
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 3383 a 2213 c 2125 g 3171 t
ORIGIN

alignment_scores:
Quality: 259.50 Length: 414
Ratio: 1.359 Gaps: 21
Percent Similarity: 46.135 Percent Identity: 25.845

alignment_block:
US-09-775-181-4 x AC017271/rev ..
Align seg 1/1 to reverse of: AC017271 from: 1 to: 10892
226 LeuArgLysTrpArg...Pro.HisLeuHisArgArgGlyProAsnG 241
|||||
9487 CTTCAGAGTTGGCGCTTCATTCGCCGCGAGGAGTGTGTC 9438
241 lnGlyProArgGlyLeuGlyHisSerTrpArgArgLysAspGlyLeuGly 257
|||||
9437 AA.....CGATGGCCGCGAGGAGTTTCGG..... 9412
258 GlyAspLysSerHisPheLysTrpSerProTrpLeuGluCysGluAs 274
|||||
9411 .....TCGCCGTCATGGGTGAGTTGGCCA 9386
274 nGlySerTyLysProGly.....TrpLeuValThrLeuSerS 287
|||||
9385 AACTTTATACGATAATTTCGTGACATTAGTTCTGGATAATA..... 9346
287 erAlaIleTyrGlyLeuGlnProAsnLeuValProGluPheArgGlyVal 303
|||||
```



```
10865 AAGTATTTAGTTTACATATCCCTAAAGTTGGGAAACTTTTGAGTAATCCCTT 10816
449 .....AlaSerGlyLeu...IleLeuLeuGlnThrI 438
10815 TCATCATTTTTCAGGCCATTCGCTGTGATGTGGAGTGTCTGCTGAGACGA 10766
458 leLeuPheGlySerLeuLeuTyPheProVal.ValIleLeu..... 472
10765 TACTGCTGGGCATGTTTACTTATTCATCTGTGAGTGTCTGCTGCGGAT 10716
472 ..... 472
10715 GTCACCTTCACACTTCGTGCTGATTACATATTCATCTTAGTTGCCG 10666
473 .....TyrPheGluProSerThrPheArgCysIleLeuLeuArgTrpAla 487
10665 TCATTTCTTTCCCGCATTCACCGAGCGTGCCTCTGGAGCCCTGGCTT 10616
488 ArgLeuLeuGlyPheAlaThrValTrpGlyThrValThrLeuLeuIleH 504
10615 CGGGAGCTGGGCTTCATCACTGCTACGGCGCATCATCACTGAAGTTGTA 10566
504 sArgValLeuLysValPheLeuSerArgThrAlaGlnArgIleProTyRM 521
10565 CCACCACCTGCTGGACTTCGCGACCCCGAAGCCATCGCTGG...GTGC 10519
521 etThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLeuVal 537
10518 TGGCGGACGTGGACCTGCTCAAGTATCTGGGCACCATGCTCTTGGCGTC 10469
538 PheTrpPheLeuIleGlyTrpThrSerSerValCysGlnAsnLeuGlu 554
10468 ATCTGCTACATGCGCCCTTTACGGCTCGTCTGTTGGATCTTCTCGAAG 10419
554 s...GlnIleSerLeuIleGlyGlnGlyLysThrSerAspHisLeuIleP 570
10418 TGGCGACGTGGAGAGCTCAGGGAGCGGACACG..... 10385
570 heAsnMetCysLeuIleAspArgTrpAspTyRMetThr 582
10384 ..AACACCTGCCATCGCTCAAGTGGGAAGTGTGTACG 10349
seq_name: gb_inl:AE003575
seq_documentation_block:
LOCUS AE003575 327209 bp DNA INV 04-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386046 section 16
of 16, complete sequence.
VERSION AE003575 AE002638
WORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 327209)
Adams,M.D., Celnikier,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.C., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Wan,K.H., Doyle,C., Rogers,Y.H., Blazer,R.G., Champe,M., Pfeiffer,B.D.,
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Frannkovich,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrieres,S.,
Fleischmann,W., Fostler,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Moberg,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Paclet,J.M., Palazolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
2 (bases 1 to 327209)
Adams,M.D., Celnikier,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7295650.
Location/Qualifiers
1. 327209
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2L"
/gene="CG11933"
/product="CT35796"
/db_xref="FLYBASE:FBan0011933"
/db_xref="FLYBASE:FBgn0031615"
/evidence=not_experimental
<29815..>30955
/gene="CG11933"
/map="25A3-25A3"
/db_xref="FLYBASE:FBan0011933"
/db_xref="FLYBASE:FBgn0031615"
/evidence=not_experimental
join(29815..29907,30773..30955)
/gene="CG11933"
/codon_start=1
/db_xref="FLYBASE:FBan0011933"
/db_xref="FLYBASE:FBgn0031615"
/evidence=not_experimental
/protein_id="AAF50997.1"
/db_xref="GI:7295691"
/translacion="MAOPFLGSPDCHLWTGLGSGCGSGSSCCDYLRHLHSPNFQD
YWQARVHREYQIFPGQMPLOYGVHDLRLFLPTNCQEYNSNQVQ"
complement(<31022..>31300)
/gene="CG11932"
/product="CT35794"
/db_xref="FLYBASE:FBan0011932"
/db_xref="FLYBASE:FBgn0031616"
/evidence=not_experimental
complement(<31022..>31300)
/gene="CG11932"
/map="25A3-25A3"
/db_xref="FLYBASE:FBan0011932"
/db_xref="FLYBASE:FBgn0031616"
/evidence=not_experimental
complement(31022..31300)
/gene="CG11932"
CDS
```


Align seg 1/1 to: AL354976 from: 1 to: 186964

333 AsnAsnSerGluCysMetProIleLysGlyLeuGlyPheValLeuGlyAl 349
: : : : :
179733 TCTTCTCATAGTGTATGCAATTAAGGCCCTAGGATCGTTCTTGGAGC 179782

349 aTyrGluCysIleCysLysAlaGlyPheTyrHisProGlyValLeuProV 366
: : : : :
179783 CTATGAGTCATTTCAAGCAGGAGTTCATCATCTCGAGTCTTACCAG 179832

366 aLAsnAsnPheArgArgGlyPro 374
: : : : :
179833 TGACAACTTCGGAGTAGTCCCT 179858

seq_name: gb_htg22:AL445244

seq_documentation_block:
LOCUS AL445244 160390 bp DNA HTG 20-JAN-2001
DEFINITION Homo sapiens chromosome 10 clone RP13-99J21, *** SEQUENCING IN
PROGRESS ***, 6 unordered pieces.
ACCESSION AL445244
VERSION AL445244.4 GI:10800726
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
Sims.S.
Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Oct 13, 2000 this sequence version replaced gi:10798428.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BB99J21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 158004 bases at least Q40
Consensus quality: 159033 bases at least Q30
Consensus quality: 159555 bases at least Q20
Insert size: 159890; sum-of-contigs
Insert size: 144185; 13.1% error; agarose-fp
Quality coverage: 5.59x in Q20 bases; sum-of-contigs Quality
coverage: 6.20x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 43907: contig of 43907 bp in length
* 43908 44007: gap of 100 bp
* 44008 55206: contig of 11199 bp in length
* 55207 55306: gap of 100 bp
* 55307 67735: contig of 12429 bp in length
* 67736 67835: gap of 100 bp
* 67836 73636: contig of 5801 bp in length
* 73637 73736: gap of 100 bp
* 73737 100153: contig of 26417 bp in length
* 100154 100253: gap of 100 bp
* 100254 160390: contig of 60137 bp in length.

FEATURES
source Location/Qualifiers
1.160390
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP13-99J21"
/clone_lib="RPC1-13.1"
1.43907
/note="assembly_fragment:01136"
clone_end:T7
vector_side:left
44008_55206
/note="assembly_fragment:01384"
fragment_chain:1
55307_67735
/note="assembly_fragment:01898"
fragment_chain:1
67836_73636
/note="assembly_fragment:01997"
fragment_chain:1
73737_100153
/note="assembly_fragment:01657"
fragment_chain:1
100254_160390
/note="assembly_fragment:01063"
fragment_chain:1
clone_end:SP6
vector_side:right
BASE COUNT 55825 a 29765 c 28364 g 45935 t 501 others
ORIGIN

alignment_scores:
Quality: 200.00 Length: 36
Ratio: 5.556 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-775-181-4 x AL445244/rev ..
Align seg 1/1 to reverse of: AL445244 from: 1 to: 160390

301 ArgGlyValMetLysValAspIleAsnLeuGlnLysValAspIleAspG1 317
|||||
136464 AGGGGTGTCATGAAAGTTGACATAATCTTCAGAAAGTGGACATTCACCA 136415
|||||

317 nCysSerSerAspGlyTTPheSerGlyThrHisLysCysHisLeuAsnA 334
|||||

136414 ATGCTCAAGTGATGCTGTTTTCAGGAACCTCAATAATGCCACCTCAACA 136365
|||||

334 snSerGlu 336
|||||

136364 ATTCAGAG 136357

seq_name: gb_htg4:AC014084

seq_documentation_block:
LOCUS AC014084 150133 bp DNA HTG 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in Ordered
pieces.
ACCESSION AC014084
VERSION AC014084.1 GI:6437251
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 150133)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA


```
240 AsnGlnGlyProArgGlyLeuGlyHisSerTrpArgArgLysaspGlyLe 256
||| |||||
45637 AACCTGCTCGC..... 45626

256 uGlyGlyAspLysSerHisPheLysTrpSerProTyrLeuGluCysG 273
|||||:|||||
45625 .....ATTAAATTACAGGCCGCTACTTGATGT. 45596

273 luAsnGlySerTyrLys.ProGlyTrp..... 281
|||:|||||
45595 .....GCAGGTCCAACAGTGGCTGGTGGCCGAGTAGTCCCAATTGCG 45551

282 .....LeuValThrLeuSer.....SerA 288
|||||:|||||
45550 GATATCTACCCCGACACAGCGAGTTCGTCACATTGAGTATCCCAAGTA 45501

288 laIleTyrGlyLeuGlnProAsnLeuValPro..... 298
|||:|||||:|||||
45500 AGATACCTTGAAATATCCCGTGAATACCTCTTTTATCTACTGATCGCT 45451

299 ...GluPheArgGlyValMetLysValAspIleAsnLeuGlnLysValAs 314
|||||:|||||
45450 TTTAGATACACGGCGGTTCGGTCTTTGAGATGACACTCGAGCGTATCGA 45401

314 pIleAspGlnCys.....SerSerAspGlyTrpPheS 325
|||||:|||||
45400 CATAAACCACTGTCCATTTGGGTGAAGCAACAAGGACCTTAATCACPTTG 45351

325 erGlyThrHisLysCysHisLeuAsnAsnSerGluCysMetProIleLys 341
|||:|||||:|||||
45350 CGGNATCGCGCGGTGTAAAGAAACGACAGAGTGTGAACCATATACAA 45301

342 GlyLeuGlyPheValLeuGlyAlaTyrGluCysIleCysLysAlaGlyPh 358
|||||:|||||
45300 GGCTGGCGCTTTAGCGCGGTGGCTACCGCTGCTGTAAGCCAGGTTT 45251

358 eTyrHisProGlyValLeuProValAsnAsnPheArgArgGlyProA 375
|||:|||||
45250 TCGCTGCCCAACGCTAGTG.....CGCGCACCTTATCTGG 45216

375 spGlnHisIleSerGlySerThrLysAspValSerGluGluAlaTyrVal 391
|||:|||||:|||||
45215 GCAGAGTTGGAGCGCGGCTCGGCAGAACAGTACTACACGAGTAGCAC 45166

392 CysLeuProCysArgGluGlyCysProPhe.....CysAlaAs 404
|||||:|||||
45165 TGCCTT.....AAGATTGCTCTATGTTTAACTAGCAATATGTAAG 45122

404 p.....AspSerProCysPheValGlnGluAspLysTyrLeuArgL 418
|||:|||||
45121 TATCAGATTTTCACTCTTGATGTTTT..... 45097

418 euAlaIleSerPheGlnGlyCysMetLeuLeuAspPheValSer 434
|||||:|||||
45096 ....TTTTAGGGATCCAAAGCTTCCCATTCAGTGGGATAAGGCTCC 45052

435 MetLeuValTyrHisPheArgLys 443
|||||:|||||
45051 .....TACCACATTCCGCCAA 45037
```

seq_name: gb_inl:AC007771

```
seq_documentation_block:
LOCUS AC007771 197071 bp DNA INV 15-FEB-2001
DEFINITION Drosophila melanogaster, Chromosome 3R, region 92E-92F, BAC clone
BACR26M03, complete sequence.
ACCESSION AC007771.7 GI:12831351
VERSION AC007771.7
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
```

REFERENCE
AUTHORS

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 197071)
Celisner, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,
Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Dou, L.E., Doyle, C., Dresnek, D., Farfan, D.,
Fieriera, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Paclet, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Shapleton, M., Strong, R., Svirskaas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

TITLE
JOURNAL
REFERENCE
AUTHORS

Sequencing of Drosophila chromosome 3R, region 92E-92F
Unpublished
2 (bases 1 to 197071)

Celisner, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Madad, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclet, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirskaas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (10-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 15, 2001 this sequence version replaced gi:7018763.
Sequence submitted by:
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bdgp@fruitfly.berkeley.edu.

FEATURES
source

1..197071
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="92E-92F"
/clone="BAC BACR26M03 (D738)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACe3.6)"
BASE COUNT 54696 a 44041 c 42886 g 55448 t
ORIGIN

alignment_scores:
Quality: 183.50 Length: 511
Ratio: 0.913 Gaps: 28
Percent Similarity: 39.335 Percent Identity: 22.701

alignment_block:

US-09-775-181-4 x AC007771/rev ..

Align seg 1/1 to reverse of: AC007771 from: 1 to: 197071

7 ProLeuLeuLeuCysLeuLeu.....LeuAlaGlnLeuGlyLeuGlyAl 21
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
171917 CCCTGCTGCATCTGCAGATATGGCCCTTAGCAGAGCTTCTTCTGGAGC 171868
21 aValGlyAlaSerArg.....AspProGlnGlyArgProAsps 34
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
171867 TACATCTCCAGTCGAGGTTTATTCGACCCGCCCATCAACGACACCTACGA 171818
34 erProArgGluArgThrProLysGlyLysProHisAlaGlnGlnProGly 50

```

171817 TCCCGCATGATGACTACTTCTCTCCACCGTAGCCCGATATCCG... 171771
51 ArgAlaSerAlaSerAspSerAlaProTrpSerArgSerThrAsp.G 67
171770 .....CCACCCACATATCAACGCCCTCG 171748
67 lyThrIleLeuAlaGlnLysLeuAlaGluGluValProMetAspValAla 83
171747 GCCGTGTAATT.....CTCCCCAACAGCTCGTATTCTGCTGTCGTA 171707
84 SerTyrLeuTyrThrGlyAspSerHisGlnLeuLysArgAlaAsn..... 98
171706 TCGCGGCTTCTTCAATAGACGTTCCCCAGATTCCGGCCAAAGACCTTCA 171657
98 ..... 98
171656 GGCTGGACGACTTCAACGATCCCATTCATCTGCAGAGATATCGACGTGG 171607
99 .....CysSerGlyArgTyrGluLeuAlaGlyLeuProGlyLys 111
171606 AATACTTTTCGATGTTTCAGGA.....TCTGGCGGCCCATCA... 171572
112 TrpProAlaLeuAlaSerAlaHisProSerLeuHisArgAlaLeuAspTh 128
171571 .....ATCCAAAG 171552
128 rLeu..ThrHisAlaThrAsnPheLeuAsnValMetLeuGlnSerAsnLys 144
171551 ACTACACCCACGACCTGTATAAATAAAC..... 171523
145 SerArgGluGlnAsnLeuGlnAspLeuAspTrpTyrGlnAlaLeuVa 161
171522 .....GAGTGTACCGCGCTGGCT 171503
161 lTrpSerLeuLeuGluGlyGluProSerIleSerArgAlaIleThrP 178
171502 ACCAGAACGTCGAGGA..... 171484
178 heSerThrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAla 194
171484 ..... 171484
195 ThrArgGluGluSerArgIleLeuLeuGlnAspLeuSerSerAlaPr 211
171483 ...CGGCACGATACGAGATCACCTACCAG.....GTGGAAT 171449
211 oHisLeuAlaAsnAlaThrLeuGluThrGluTrpPheHisGlyLeuArgA 228
171448 CCGCTATCGGAACAACAACACGAGAGCTATACCTTCCAC..... 171409
228 rGlySerTrpArgProHisLeuHisArgArgGlyPro..... 239
171408 .....GACCGCTCGCTCTCAAGAA 171388
240 AsnGlnGlyProArgGlyLeuGlyHisSerTrpArgArgLysAspGlyLe 256
171387 AACCTGTGTCGG..... 171376
256 uGlyGlyAspLysSerHisPheLysTrpSerProProTyrLeuGluCysG 273
171375 .....ATTAAATTTACAAGCCGCTACTTCGATGTT. 171346
273 luAsnGlySerTyrLys..ProGlyTrp..... 281
171345 .....GGCAGTCCAACAAGTGGCTGGCGCCGACGATAGTCCCAATGCCG 171301
282 .....LeuValThrLeuSer.....Sera 288
171300 GATATCTACCCCGACACACGACGAGTTCGTCACATTTGAGTATCCCAAGTA 171251
288 lalleTyrGlyLeuGlnProAsnLeuValPro..... 298

```

```

171250 AGATACCTTGAATATCCCTCAATACCTCTCTTTTATCTACTGTATCGCT 171201
299 ...GluPheArgGlyValMetLysValAspIleAsnLeuGlnLysValas 314
171200 TTTAGATACACGCCGCTTTCGTTCTTGAGATGGACTTCGAGCGTATCGA 171151
314 pIleAspGlnCys.....SerSerAspGlyTrpPheS 325
171150 CATAAACCAAGTGTCCATTTGGGTGAAGCGACAAAGACCTAATCACTTTG 171101
325 erGlyThrHisLysCysHisLeuAsnAsnSerGluCysMetProFileLys 341
171100 CGGATACGCCCGGTGTAAAAAAGAACACAGAGTGTGAACCAATTACAA 171051
342 GlyLeuGlyPheValLeuGlyAlaTyrGluCysIleCysLysAlaGlyPh 358
171050 GGCTGGGCTTTAGCGCGGTGGCTACCAGTCCGCTTGTAAAGCCAGGTTT 171001
358 eTyrHisProGlyValLeuProValAsnAsnPheArgArgGlyProA 375
171000 TCGGCTGCCCAACGTAGTG.....CGCGACCTTATCTGG 170966
375 spGlnHisIleSerGlySerThrLysAspValSerGluGluAlaTyrVal 391
170965 CGGAGATTGTGGAGCGGCATCGCGCAACAGACTACTACAAGAGTACGAC 170916
392 CysLeuProCysArgGluGlyCysProPhe.....CysAlaAs 404
170915 TGCCTT.....AAGATTGGCTGTATGTTTAAAGTAGCAATATGTANAAG 170872
404 p.....AspSerProCysPheValGlnGluAspLysTyrLeuArgL 418
170871 TATGAGATTTGACTCTTGTATGTTT..... 170847
418 euAlaIleLeSerPheGlnGlyLeuCysMetLeuLeuAspPheValSer 434
170846 .....TTTTAGGGATCCAAAGCTTCCCATTCAGTGGGATAAGGCCCTCC 170802
435 MetLeuValValTyrHisPheArgLys 443
170801 .....TACCACATTCGCCAA 170787

```

seq_name: gb_in2:AE003732

seq_documentation_block:

LOCUS AE003732 234369 bp DNA INV 05-OCT-2000
 DEFINITION Drosophila melanogaster genomic scaffold 142000013386035 section 57
 of 105, complete sequence.

ACCESSION AE003732 AE002708

VERSION AE003732.1 GI:7300654

KEYWORDS HTG.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 234369)

AUTHORS

Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
 Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
 George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
 Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
 Brandon,R.C., Rogers,Y.H., Blazer,R.G., Champe,M., Pfeiffer,B.D.,
 Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
 Miklos,G.L., Abril,J.F., Aghayani,A., An,H.J., Basu,A.,
 Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Beeson,K.Y.,
 Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
 Benos,P.V., Bertram,B.P., Bhandari,D., Bolshakov,S., Borokova,D.,
 Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
 Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
 Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
 Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
 Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
 Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrieria,S.,


```
DDTDSVRDLRRRIKLEHITVSIIEEIRQYRHILEQHITTLNMLHDFMREKNRIYSD
FAKSSYILFRKFNSTOKNILLKKLELTCDDKSSPHITTMALRELLRELQKRSKDVQNC
ATLLIPMLDRISLSTQTVAMDLCHVAFDPNLSQCLQEQVDMVVKKOLINSI
DNFKQGIIGCVGLIDAMARIANNVDRDEFIASVENVDSLDPGRGKMAANLIIRTEA
SIGNSTESIALPFEELATFNORNECTSGCELDNQFIWACDLVTRFOASFVTENVP
ETLKACDSLVLAPLENYVVLVKKRHODSLESINALLGCAIVLPSEFDDNNYVSVE
NFEAEQKDLISYFIHVNWMRVSISAFASQRPPTRRRVLRLGELIRIEQRMKPLL
ARAPVDFAPPYQFLNVLSNQNRKPGPKPAKLNATLPEPDLTGNOPSIADFTIK
VGCKTVTKTDEQMYGPRERYRMEVEIIMLVQKFNHQLHEEQMGFGLGLE
LRFLLBEDVVKLEAAVLRHDSYDADFRLAKPDEFICDILPCLHVEVNNHLITLGE
AIDONLTVSHVYSLNLDKQDFCIKSCFGLCVRLFALVFAMSEWSDKSOEOLLHRI
LCGTLRLRWKWHVSGTLDKQOCNIYLDLVKGLKSNKQSOTELLTLYKQCSILN
TKDKALITSPNFKANPPLLFRGLCEVLHLSGQVSDSRGKLSWASVDLLNGL
LSIVQVEPRNGLKHLKSLFLKLLQHGMSALESIVREDPERLRLHLELQKVTIR
FLQLCCHSKSIKNTAISYIPSLRETIETLVRKALLAANNCHSAFHMGNMNRDL
HGDSIITPRSSPFAGENSDDELPAADTSDVETVLGDDMGITAVSVTRPSDGRSRKKS
SSRSKCF"
mRNA
complement(join(67952..68088,68183..68301,68369..68459,
68514..68667,69583..>69798))
/gene="CG17270"
/product="CTJ35903"
/db_xref="FLYBASE:FBan0017270"
/db_xref="FLYBASE:FBgn0038828"
complement(<67952..>69798)
/gene="CG17270"
/map="92F11-92F13"
```

alignment_scores:

```
Quality: 183.50      Length: 511
Ratio: 0.913        Gaps: 28
Percent Similarity: 39.335 Percent Identity: 22.701
```

alignment_block:

US-09-775-181-4 x AE003732/rev ..

Align seg 1/1 to reverse of: AE003732 from: 1 to: 234369

```
7  ProLeuLeuLeuCysLeu.....LeuAlaGlnLeuGlyLeuGlyAl 21
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12771 CCTGCTGCATCTGCAGATATAGGCCCTTAGCAGAAGCTTCTCTGGAGC 12722

21  aValGlyAlaSerArg.....AspProGlnGlyArgProAspS 34
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12721 TACATCTCCAGTCGAGTGTATTTCGCCCGCCATCAACGACACCTACGA 12672

34  erProArgGluArgThrProLysGlyLysProHisAlaGlnGlnProGly 50
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12671 TCCCGGCATGATGACTACTTCTCTCCACCGTAGCCGATGATCCG... 12625

51  ArgAlaSerAlaSerAspSerSerAlaProTrpSerArgSerThrAsp.G 67
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12624 .....CCAAACCCACATATCAACGCCCTCG 12602

67  lyThrIleLeuAlaGlnLysLeuAlaGluGluValProMetAspValAla 83
   ||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12601 GCGGTGACTT.....CTCCCCAACAGCTCGTATTCGTCTCGTA 12561

84  SerTyLeuTyThrGlyAspSerHisGlnLeuLysArgAlaAsn..... 98
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12560 TCGCGGCTTCTTCAATAGACGTTCCTCCAGATTCCGGCCAGAACCTTCA 12511

98  ..... 98

12510 GGCTGGACGACTTCAACGATCCCATTCATCTGCAGAAGATATCGACGTGG 12461

99  .....CysSerGlyArgTyArgGluLeuAlaGlyLeuProGlyLys 111
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12460 AATACTTTTCGATGTTTCAGGA.....TCTGGGCGCCCATCA... 12426

112  TrpProAlaLeuAlaSerAlaHisProSerLeuHisArgAlaLeuAspTh 128
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12425 .....CCCAGACTCCAT.....ATCCAAGG 12406
```

```
138 rLeu.ThrHisAlaThrAsnPheLeuAsnValMetLeuGlnSerAsnLys 144
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12405 ACTACACCCACGACCTGTATAAAATAAAC..... 12377

145  SerArgGluGlnAsnLeuGlnAspLeuAspLeuAspTrpTyGlnAlaLeuVa 161
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12376 .....GAGTGGTACCGCGCTGGCT 12357

161  lTrpSerLeuLeuGluGlyGluProSerIleSerArgAlaAlaIleThrP 178
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12356 ACCAGACAACGTCGAGGA..... 12338

178  heSerThrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAla 194
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12338 ..... 12338

195  ThrArgGluGluSerArgIleLeuLeuGlnAspLeuSerSerSerAlaPr 211
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12337 ...CGGCACGATACGACGATCACCTACACG.....GTGGAAT 12303

211  oHisLeuAlaAsnAlaThrLeuGluThrGluTrpPheHisGlyLeuArga 228
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12302 CGCTATTCGGAACAACAACAGACGCTATACCTTCCAC..... 12263

228  rgLysTrpArgProHisLeuHisArgArgGlyPro..... 239
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12262 .....GGACCGCTGGCTCTGAAGAA 12242

240  AsnGlnGlyProArgGlyLeuGlyHisSerTrpArgArgLysAspGlyLe 256
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12241 AACCTGTGTCG..... 12230

256  uGlyGlyAspLysSerHisLysTrpSerProProTyLeuGluCysG 273
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12229 .....ATTAAATTTACAAGCGCGTACTCTCGATTGT. 12200

273  luAsnGlySerTyLys.ProGlyTrp..... 281
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12199 .....GGCAGGTCCAACAAGTGGCTGGTGGCGCAGTAGTGCAATTGCG 12155

282  .....LeuValThrLeuSer.....SerA 288
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12154 GATATCTACCCCGACACACGCGTTCGTCACATGATGATGCCAAGTA 12105

288  laIleTyGlyLeuGlnProAsnLeuValPro..... 298
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12104 AGATACCTTGAATATCCCTGAATACCCCTCTTTTATCTACTGTATCGCT 12055

299  ...GluPheArgGlyValMetLysValAspIleAsnLeuGlnLysValas 314
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12054 TTTAGATACACGCGCGTTCGTTCTTGAGATGGACTTCGACCGGTATCGA 12005

314  pileaspGlnCys.....SerSerAspGlyTrpPheS 325
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
12004 CATAACCAAGTGTCCATTTGGGTGAAGGCAACAAGACCTAATCACTTTG 11955

325  erGlyThrHisLysCysHisLeuAsnSerGluCysMetProIleLys 341
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
11954 CGGATACGCGCGGTGTAAAAAAGAAACACAGAGTGTGAACCATTAACA 11905

342  GlyLeuGlyPheValLeuGlyAlaTyArgGluCysIleCysLysAlaGlyPh 358
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
11904 GGCTGGGCTTTAGCGCGGTGGCTTACCAGTCCCGTTCGAGCCAGGTTT 11855

358  eTyHisProGlyValLeuProValAsnAsnPheArgArgArgGlyProA 375
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
11854 TCGGCTGCCCAACGCTAGT.....CGGCGACCTTATCTGG 11820

375  spGlnHisIleSerGlySerThrLysAspValSerGluGluAlaTyVal 391
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
11819 CGGAGATTGGAGCGCGCATCGGCGAGAACAGTACTACACGAGTAGCAC 11770

392  CysLeuProCysArgGluGlyCysProPhe.....CysAlaAs 404
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

```
|||||
11769 TGCCTT.....ANGATTGCTGTATGTTTAAAGTAGCAATATGTAAAG 11726
:
404 p.....AspSerProCysPheValGlnGluAspLysTyrLeuArgL 418
: ||||| |||||
11725 TATGAGATTGACTCTTGATGTTT..... 11701
: ||||| |||||
418 euAlaIleSerPheGlnGlyLeuCysMetLeuLeuAspPheValSer 434
: ||||| ||||| ||||| |||||
11700 .....TTTTAGGATCCAAAGCTTCCCATTCAGTGGGATAAGGCTCC 11656

435 MetLeuValTyrHisPheArgLys 443
|||||
11655 .....TACCACATTCGCCAA 11641
```

seq_name: gb_in3:DMX99675

```
seq_documentation_block: 3325 bp mRNA INV 07-FEB-1997
LOCUS DMX99675
DEFINITION D.melanogaster mRNA for metabotropic glutamate receptor.
```

X99675

X99675.1 GI:1834426

metabotropic glutamate receptor.

fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 3325)

Parmentier, M.L., Pin, J.P., Bockaert, J. and Grau, Y.

Cloning and functional expression of a Drosophila metabotropic

glutamate receptor expressed in the embryonic CNS

J. Neurosci. 16 (21), 6687-6694, (1996)

96421661

2 (bases 1 to 3325)

Parmentier, M.

Direct Submission

Submitted (31-JUL-1996) M. Parmentier, CCIPE-UPR 9023, Upr 9023,

141 Rue De La Cardonille, 34094 Montpellier Cedex 05, F-34094,

FRANCE

Location/Qualifiers

1. 3325

/organism="Drosophila melanogaster"

/strain="Oregon R"

/db_xref="taxon:7227"

/chromosome="4"

/dev_stage="adult"

190..3120

/codon_start=1

/product="metabotropic glutamate receptor"

/protein_id="CAA67993.1"

/db_xref="GI:1834427"

/db_xref="FLYBASE:FBgn0019985"

/db_xref="SWISS-PROT:P91685"

```
/translation="MKQKNNGTILVVMVLSWRVVDLKSPTHQDSVSVSLPGD
IILGLPFVHEKGEAGPCGPKVYNRGVORLEAMLYADRVNNDNPILGHTIGVHID
TCSRDTVALNQSLQFVRSALNNLTSGYECADGSSPOLRKNASSCPVFGVIGSYSSV
SLOVANLLRLFHIPQVSPASTAKTLSDKTFDLFARTVPDTPFOSVALVDILKNFNS
YVSTIHSEGSVEGYEIALHKEATERNVCIATAEKVPSAADKVFDSILSKLQKNPA
RGVLFTRAEADRILOAKRANLSOPHMTASDGKQKQLLEGLEDIARGATTVEL
QSEIIADFDYRMQLTETQNRNFWFAEYEDTENCVLTSVAPDYSNSANSTDNKL
GKAKTECDSDYSRUSEKVGESKTQFVVDVAFAYALHNLHNDRCNTQSDQTTET
RVKLQESVWYRKISTDTKSOACPDMDNYGKEYNNYLLNVSPIDLAGEVRFDRQG
DLARYDILLYNQRENSGQYKVIQKFWGLQNLSETVVMNKEQTPSACSLPCSV
GMIKKQDQTCCTCDSESEFYVYDEFTCKDCGGLWPYADKLSCYALDIQYMKWS
LFAIPALIAFGALTSTIVLFKKNIDPLVRASGRELSYTLFGILCYCNPTEL
IAKTSICSVLQRFQGVGSFIYSALLTYNRISRHSASQAQRLKLYCSPQSV
CTIYAIKRIKPIENFNSKTEGFTMYTKTLIWLAFAPYFTGNSYEVQVTTLCIS
LSASVALVCLYSPKVIYILVFPDKNRKLTMINSTVYRRSAAVAQAQPTSGSYRTHA
PGTSALTGGAGVTNASSSTLPTQNSPHLDEASQTNVAHNTNGEFLPEVGERVEPICH
IVNK"
```

BASE COUNT 1038 a 680 c 687 g 920 t

.. . .

ORIGIN

alignment_scores:

Quality: 164.00 Length: 494

Ratio: 0.710 Gaps: 24

Percent Similarity: 46.761 Percent Identity: 20.648

alignment_block:

US-09-775-181-4 x DMX99675

Align seg 1/1 to: DMX99675 from: 1 to: 3325

```
145 SerArgGluGlnAsnLeuGlnAspLeuAspTyrTrpGlnAlaLeuVa 161
: : : : : ||||| : : : : :
1072 GCTAAACGAGCTTAACATTATCGCAACCACTTTCATTGGATGCAAGTGTGG 1121
: : : : : ||||| : : : : :
161 lTrp.....SerLeuLeuGluGlyGluProSerIleSerArgA 174
: : : : : ||||| : : : : :
1122 TTGGGGTAAACAAACAACTACTCGAAGGCTTAGAGGACATCGCTGAGG 1171
: : : : : ||||| : : : : :
174 laAlaIleThrPheSerThrAspSer...LeuSerAlaProAlaProGln 189
: ||||| ||||| : : : : :
1172 GTGCAATACTGTAGAACTGCGAGTCAGATATATCGACATTTTGATCGC 1221
: : : : : ||||| : : : : :
190 ValPheLeuGlnAlaThrArgGlu.Glu..... 198
: : : : : ||||| : : : : :
1222 TACATGATCAACTGACGCCAGACAGACAACTCAAAGCAATCCTTTGGTTGC 1271
: : : : : ||||| : : : : :
198 ..... 198
1272 TGAATACTGGGAAGATACATCAACTGTGTTTGGACATCTCTTTTCGGTAA 1321
199 .....SerArgIleLeuGlnGlnAspLeuSerSerAl 210
: : : : : ||||| : : : : :
1322 AACCTGATACTTCAAACTCGGCTAATTCTACTGACAAATAAATCGCGGTC 1371
: : : : : ||||| : : : : :
210 aProHisLeuAlaAsnAlaThrLeuGluThrGluTrpPhe..... 223
: : : : : ||||| : : : : :
1372 AAAGCAAAACAGAAATCGCAGCATGTTACCGGTTTATCGAAAAAGTTGG 1421
: : : : : ||||| : : : : :
224 .....HisGlyLeuArgArgLysTrpArgProHis 233
: : : : : ||||| : : : : :
1422 ATACGAACAGGAGTCAAGACACAAATTG.....TTGTGGATGCGCGTCT 1465
: : : : : ||||| : : : : :
234 LeuHis..... 235
: : : : : ||||| : : : : :
1466 AGCATTTGCTTATGCATTCATAATTTCATAATGATGTTGCAACACA 1515
: : : : : ||||| : : : : :
236 .....ArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHis..Se 249
: : : : : ||||| : : : : :
1516 CAAAGCGACCAAGCACTGACAGCAAGCAAGCACTACAAGCAATCTGT 1565
: : : : : ||||| : : : : :
249 rTrpArgArgLysAspGlyLeuGlyClyAspLysSerHisPheLysTrpS 266
: : : : : ||||| : : : : :
1566 GTGGTATCGAAAA.....ATTTCGACTGAT...ACTAAGTCCCAAGCGT 1606
: : : : : ||||| : : : : :
266 exProProTyrLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeu 282
: : : : : ||||| : : : : :
1607 GTCTGCACATGCTAATTATGCGGAAAGAGGTTTTCACAACTTACCTT 1656
: : : : : ||||| : : : : :
283 ValThrLeuSerSera...lIeTyrGlyLeuGlnProAsnLeuValPr 298
: : : : : ||||| : : : : :
1657 TTGAAGCTATTCATACATCTTCTGTCGCGCAAGGTTAAATTCGATCG 1706
: : : : : ||||| : : : : :
298 oGluPheArgGlyValMetLysValAspile...AsnLeuGlnLysValA 314
: : : : : ||||| : : : : :
1707 ACAGGCGGATGTTGGCCAGATACGACATTTTGAATTATCAACCG.... 1752
: : : : : ||||| : : : : :
314 spIleAspGlnCysSerSerAspGly..... 322
: : : : : ||||| : : : : :
1753 .....CAAGAAACTCTTCAGGATATCAGTATAAGGTTATTGGAAAA 1794
```

```
323 TrpPheSerGlyThrHisLysCysHisLeuAsnSerGluCysMetPr 339
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1795 TGGTTTAATGGT.....TTGCAACTAAACTCAGAAACAGTTGTATG 1835

339 oileLysGlyLeuGlyPheValLeuGlyAlaTyrGluCysLysA 356
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1836 GAACAAGGAACCGAACCACTACTTCGGCCCTGTTCACTACCATGTGAGG 1885

356 laGlyPheTyrHisProGlyValLeuProValAsnAsnPheArgArg 372
:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:
1886 TCGGAATG.....ATCAAGAAACAA 1905

373 GlyProAspGlnHisIle.....SerGlySerThrLysAspVa 385
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1906 CAGGGGATACCTGTGTGGATATCGACAGCTGTGAATCTTTGAATA 1955

385 lSerGluGluAlaTyrValCysLeuProCysArgGluGlyCysProPheC 402
:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:
1956 CGTATATGATGAGTTAGGTGTAAGACTGCGGCTCGACTTTGGCCCT 2005

402 ysAlaAspSerProCysPheValGlnGluAspLysTyrLeuArg... 417
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2006 ATGCTGACAAGCTCTCGCTATGCTTTAGACATCCAGTATATGAATGG 2055

418 .....LeuAlaIleIleSerPheGlnGlyLe 426
:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:
2056 AACTCGTTGTTTGCCTTATTCGGATGGCAATTGCAATTTT...GGTAT 2102

426 uCysMetLeuLeuAspPheValSerMetLeuValValTyrHisPheArgL 443
:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:
2103 TGCACGTG.....ACCAGTATGTATAGTTTATTATTGGCAA 2137

443 ys.....AlaLysSerIleArgAlaSerGlyLeuIleLeuLeuGlu 456
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2138 AAAATCATGACACTCCATTGGTAAGAGCATCAGGACGAGCTAAGTTAT 2187

457 ThrIleLeuPheGlySerLeuLeuTyrPheProValValIleLeuTy 473
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2188 ACTCTCTTTTGGTATACTGGTTGTATTGTAATACTTTTGCACGTAT 2237

473 rPheGluProSerThrPheArgCysIleLeuLeuArgTrpAlaArgLeuL 490
:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:
2238 ACCTAAGCCACAANTGGCTCTGTGTTCTACAAGGTTTGGTATGGGG 2287

490 euGlyPheAlaThrValTyrGlyThrValThrLeuLysLeuHisArgVal 506
:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:
2288 TTGGCTTTTCCATTATATACAGTGCCTTCTAACAAACAAATCGAATT 2337

507 LeuLysValPhe.....LeuSerArgThrAlaGlnArgIleProTyrMe 521
:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:
2338 TCACGAATCTTTCACTCTGCTTCTAAGTCAGCACAGCGCCTTAAGTATAT 2387

521 tThrGlyGlyArg.....ValMetArgMetLeuAlaVal....I 533
:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:
2388 TAGTCCCAATCACAGGTGGTATACCTCTCTGATCGCAATACAGG 2437

533 leLeuLeuValValPheTrpPheLeuIle 542
:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:
2438 TGCTATCACATGATTTGGATGGTAGTT 2466
```


OM of: US-09-775-181-4 to: N_Geneseq_0601.* out_format : pfs
 Date: Nov 1, 2001 11:04 AM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODE=frame-p2n.model -DEV=xlh
-O/cgn2_1/USPTO_spool/US09775181/runat_30102001_161056_13847/app_query.fasta_1.1955
-DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -CGAPEXT=0.050 -YGAPOP=10.000 -YGAPEXT=0.500
-YGAPOP=6.000 -YGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS-human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09775181_@CEN1_1_509 -NCPU=6 -ICPU=3 -LONGLOG
DEV_TIME=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-775-181-4
Query length: 599
Database: N_Geneseq_0601.*
Database sequences: 730101
Database length: 313950809
Search time (sec): 192.710000
```

score_list:

Sequence	Strd Orig	zScore	Escore Len	Documentation
/SIDS1/gcgdata/geneseq/NA2000.DAT:AAA41583		260.00	377.96	4.4e-13
/SIDS1/gcgdata/geneseq/NA1997.DAT:AAZ31057		173.50	223.51	0.0002
/SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ31058		173.50	219.90	0.0003
/SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ30616		173.50	222.13	0.0002
/SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ30615		171.50	219.67	0.0003
/SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ30922		161.00	203.71	0.0022
/SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ31049		161.00	203.71	0.0022
/SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ31048		161.00	203.71	0.0022
/SIDS1/gcgdata/geneseq/NA1996.DAT:AAZ31382		161.00	201.92	0.0028
/SIDS1/gcgdata/geneseq/NA1998.DAT:AAZ30964		161.00	201.92	0.0028
/SIDS1/gcgdata/geneseq/NA1998.DAT:AAZ30959		161.00	201.92	0.0028
/SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ32055		161.00	201.92	0.0028
/SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ32485		161.00	201.92	0.0028
/SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ30928		161.00	201.92	0.0028
/SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ30617		158.50	202.06	0.0028
/SIDS1/gcgdata/geneseq/NA1998.DAT:AAZ30962		152.50	186.11	0.0214
/SIDS1/gcgdata/geneseq/NA1998.DAT:AAZ30958		152.50	186.11	0.0214
/SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ32053		152.50	186.11	0.0214
/SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ32483		152.50	186.11	0.0214
/SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ30926		152.50	186.11	0.0214
/SIDS1/gcgdata/geneseq/NA1996.DAT:AAZ31381		150.00	183.07	0.0316
/SIDS1/gcgdata/geneseq/NA1998.DAT:AAZ30963		150.00	183.07	0.0316
/SIDS1/gcgdata/geneseq/NA1998.DAT:AAZ30958		150.00	183.07	0.0316
/SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ32054		150.00	183.07	0.0316
/SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ32484		150.00	183.07	0.0316
/SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ30927		150.00	183.07	0.0316
/SIDS1/gcgdata/geneseq/NA1997.DAT:AAZ30916		148.50	183.35	0.0305
/SIDS1/gcgdata/geneseq/NA1998.DAT:AAZ30965		143.50	175.78	0.0805
/SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ32056		143.50	175.78	0.0805
/SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ32486		143.50	175.78	0.0805
/SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ30929		143.50	175.78	0.0805
/SIDS1/gcgdata/geneseq/NA1997.DAT:AAZ30920		139.50	170.01	0.1688
/SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ31435		135.00	127.00	41.99
/SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ30987		134.00	164.16	0.3573
/SIDS1/gcgdata/geneseq/NA1995.DAT:AAQ80418		131.00	158.32	0.7564
/SIDS1/gcgdata/geneseq/NA1997.DAT:AAZ30915		130.00	158.47	0.7413
/SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ31451		129.50	128.76	33.51
/SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ30743		127.50	157.05	0.8904
/SIDS1/gcgdata/geneseq/NA1992.DAT:AAQ25815		126.50	148.66	2.61

```
/SIDS1/gcgdata/geneseq/NA2000.DAT:AAA59555 + 125.50 154.71 1.20
/SIDS1/gcgdata/geneseq/NA1997.DAT:AAZ30616 + 123.50 149.64 2.30
/SIDS1/gcgdata/geneseq/NA1992.DAT:AAQ25812 + 122.50 145.04 4.15
/SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ31055 + 120.00 146.85 3.29
/SIDS1/gcgdata/geneseq/NA1996.DAT:AAZ16710 + 120.00 146.84 3.30
```

seq_name: /SIDS1/gcgdata/geneseq/NA2000.DAT:AAA41583

seq_documentation_block:

```
ID AAA41583 standard; cDNA; 182 BP.
XX
AC AAA41583;
XX
DT 21-AUG-2000 (first entry)
XX
DE Human secreted expressed sequence tag SEQ ID NO:323.
XX
XX Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;
XX expressed sequence tag; EST; probe; chemotactic; proliferative;
XX immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
XX thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
XX antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;
XX antitumor; osteoprotective; neuroprotective; nootropic; antipsoriatic;
XX cerebroprotective; anticonvulsant; antidepressant; gene therapy;
XX vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
XX central nervous system disorder; Alzheimer's disease; stroke;
XX Parkinson's disease; Huntington's disease; coagulation disorder;
XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
XX tumour; infection; depression; psoriasis; ss.
XX
XX Homo sapiens.
XX OS
XX WO2000021990-A1.
XX
XX 20-APR-2000.
XX
XX 15-OCT-1999; 99WO-US24205.
XX
XX 15-OCT-1998; 98US-0104435.
XX
XX (GEM) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Morberg D, Treacy M;
XX
XX WPT: 2000-317937/27.
XX
XX Isolated polynucleotides, and encoded proteins, comprising secreted
XX expressed sequence tags (SESTs), useful for treating various disorders
XX such as autoimmune, infectious, and central nervous system disorders -
XX
XX Claim 1; Page 243; 618pp; English.
XX
XX AAA41261 to AAA43419 represent specifically claimed secreted expressed
XX sequence tags (SESTs), isolated from human, mouse, xenopus and rat
XX tissue sources. The SESTs can have a range of activities depending on
XX the tissues they were isolated from. The activities include:
XX chemotactic; proliferative; immunomodulatory; haematopoietic;
XX chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
XX cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
XX antiasthmatic; vulnary; antitumor; osteopathic; neuroprotective;
XX nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
XX anticonvulsant; and antidepressant. The SESTs can be used for gene
XX therapy and in vaccines. The SESTs are useful as probes for the
XX identification and isolation of full-length cDNAs and genomic DNA
XX molecules which correspond to the SESTs. Proteins encoded by the SESTs
XX are useful in assays for determining biological activity and raising
XX antibodies. They may be useful for treatment of autoimmune disorders
XX (multiple sclerosis, insulin dependent diabetes), allergic conditions
XX (asthma), myeloid or lymphoid cell deficiencies, wounds, ulcers,
XX osteoporosis, osteoarthritis, central nervous system disorders
```



```
|||||
1339 GCCTCTGTCGACAGATCAAGAAAGTTGAGCGCTGGCAG.....GT 1379
364 uProValAsnAsnPheArgArgArgGlyProAspGlnHisIleSerGlyS 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1380 GCCCTTCTCAACTGCACCGCA.....GACTGCCTGGCAGGA 1417
381 erThrLysAspValSerGluGlu.....AlaTyrValCysLeu 393
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1418 CCAGAAAGGATCATTTGAGGGGAGCCACCTGCTGCTTTGAGTGTG 1467
394 ProCysArgGluGly.....CysPr 400
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1468 GAGTGTCTGATGGGAGATAGTAGTGAGACAGATGCCAGTGCCTGTAA 1517
400 oPheCysAlaAspAsp.....SerProCysPheV 410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1518 CAAGTGGCCAGATGACATCTGCTGTCANTGAGACACACCTCTGCAATG 1567
410 aGlnGluAspLysTyrLeuArg.....LeuAlaIle 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1568 CCAAGGAGATCGAGTTCTGCTGTCGAGGAGCCCTTTGGGATCGCACTC 1617
421 IleSerPheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeuVa 437
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1618 ACCCTCTTTCGCGCTGGGCATTTCTGACAGCCCTTTGCTGGTGT 1667
437 lValTyrHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuIleL 454
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1668 GTTATCAAGTTCGCAACACACCACTTGTCAAGGCCACCAACCGAGGC 1717
454 euLeuGluThrIleLeuPheGlySerLeuLeuTyrPheProValVal 470
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1718 TCTCTACCTCCTCTCTCTGCTGCTGCTGCTCTCTCTCTCTCTCTCT 1767
471 IleLeuTyrPheGluProSerThrPheArgCysIleLeuLeuArgTrpAl 487
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1768 TTCTTCATCGGGAGAGCCCAAGACTGACCTGCGCTGCGCCAGCCGGC 1817
487 aArgLeuGlyPheAlaThrValTyrGlyThrValThrLysLeuH 504
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1818 CTTTGGCATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1867
504 isArgValLeuLysValPheLeuSerArgThrAlaGlnArgIleProTyr 520
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1868 ACCGTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1917
521 MetThrGlyArgValMetArgMetLeuAlaVal..... 532
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1918 AAGTGTGGGGCTCAACCTGCAGTTCCTGCTGCTGCTGCTGCTGCTGCT 1967
533 .....IleLeuLeuValValPheTrpPheLeuIleGlyTrpThrSers 547
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1968 CATGCAGATTTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2017
547 erVal...CysGlnAsnLeuGluLysGlnIleSerLeuIle..... 559
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2018 GCTACCCGAACCAAGAGGATGAGATGATCATCTTCATCAGGTGCCAC 2067
560 ...GlyGlnGlyLysThrSerAspHisLeuIlePheAsnMetCysLeuI 575
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2068 GAGGGTCCCTCATGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2117
575 e 575
2118 G 2118
```

seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAx90922

seq_documentation_block:

ID AAx90922 standard; DNA; 3234 BP.

XX

AC AAx90922;

```
XX 17-JAN-2000 (first entry)
DT
XX Human calcium receptor nucleic acid sequence.
DE
XX Human calcium receptor; GABAB modulator; GABAB receptor protein;
KW gamma-aminobutyric acid receptor; metabotropic receptor;
KW chimeric protein; synaptic transmission; extracellular binding domain;
KW antagonist activity; intracellular domain; intracellular effect; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..3234
FT /tag= a
FT /product= "Human calcium receptor protein"
XX
PN WO9951636-A2.
XX
PD 14-OCT-1999.
XX
PF 02-APR-1999; 99WO-US07352.
XX
PR 03-APR-1998; 98US-0080676.
XX
PA (NPSp-) NPS PHARM INC.
XX
PI Garrett JE, Simin RT, Busby JG, Stormann TM;
XX
DR WPI; 1999-610994/52.
DR P-PSDB; AAY28840.
XX
PT Novel nucleic acids, used to screen for specific modulators, e.g. for
PT treating spasticity or Alzheimer's disease.
XX
PS Disclosure; Fig 3A-3D; 78pp; English.
XX
CC The present sequence is the human calcium receptor nucleic acid sequence.
CC This is used in assaying for compounds modulating GABAB activity. GABAB
CC receptors are metabotropic receptors that modulate synaptic transmission
CC in brain. Portions of calcium receptor are swapped with portions of
CC GABABR2 to create protein fusions. The agonist extracellular binding
CC domain is swapped for measurement of antagonist activity. Intracellular
CC domains are swapped for measuring intracellular effects caused by the
CC different receptor.
XX
SQ Sequence 3234 BP; 739 A; 931 C; 829 G; 735 T; 0 other;
```

alignment_scores:

```
Quality: 161.00 Length: 676
Ratio: 0.583 Gaps: 27
Percent Similarity: 40.828 Percent Identity: 20.118
```

alignment_block:

US-09-775-181-4 x AAx90922

Align seg 1/1 to: AAx90922 from: 1 to: 3234

```
75 AlaGluCluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
394 TCAGACACATTCCTCTCTACGATTGCTGGTGGGAGCAACTGCCTCAGG 443
91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
444 CGTCTCCAGCGCAGTGGCAAACTCTGCTGGGGCTCTCTACATT..... 486
108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
487 .....CCCGAGTCAGTTATGCCCTCTCCACG.....AGA 516
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG 141
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

[illegible]

568 uillePheAsnMetCysLeuIle 575
|||||
2328 GATCGCTACACCTGCTGCTG 2349

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ31049

seq_documentation_block:

ID AAZ31049 standard; cDNA; 3234 BP.

AC AAZ31049;

DT 07-JAN-2000 (first entry)

XX Human Car coding sequence.

XX G-protein fusion receptor; Car; calcium receptor; GluR; head injury;
KW metabotropic glutamate receptor; GABABR; chimeric receptor; stroke;
KW gamma-aminobutyric acid receptor; allosteric modulator; antagonist;
KW spinal cord injury; epilepsy; ischaemia; hypoglycaemia; anoxia;
KW Alzheimer's disease; hyperparathyroidism; osteoporosis; depression;
cognitive disorder; ss.

US Homo sapiens.

XX WO9951641-A1.

PN 14-OCT-1999.

XX 02-APR-1999; 99WO-US07333.

XX 03-APR-1998; 98US-0080671.

XX (NPSF-) NPS PHARM INC.

XX Stormann TM, Hammerland LG, Storjohann LL, Busby JG, Garrett JE;
PI Simin RT;

XX WPI; 1999-610995/52.

DR P-PSDB; AAY49105, AAY49110, AAY49115.

XX New G-protein fusion receptors and chimeras containing domains from
PT different receptors, used to screen for modulators, potentially useful
PT e.g. for treating or preventing stroke or Alzheimer's disease

XX Disclosure: Fig 5; 255pp; English.

XX The invention relates to G-protein fusion receptors (I) comprising:
CC (1) in the N to C direction, extracellular (ECD), transmembrane (TMD) and
CC intracellular (ICD) domains, each chosen independently from a Car
CC (calcium receptor), GluR (metabotropic glutamate receptor) and GABABR
CC (gamma-aminobutyric acid receptor); (2) an optional linker attached to
CC the C-terminus of ICD; and (3) a G-protein (GP) linked to ICD or the
CC linker. (I), and recombinant chimeric receptors (CR) without the GP
CC component, are used to assess function of the various domains and to
CC identify compounds (e.g. allosteric modulators or antagonists) that act
CC on these domains. The modulators are potentially useful for treating or
CC preventing diseases associated with the receptors, e.g. stroke, head or
CC spinal cord injury, epilepsy, ischaemia, hypoglycaemia, anoxia,
CC Alzheimer's disease, hyperparathyroidism, osteoporosis, cognitive
CC disorders and depression. Nucleic acid (II) that encodes (I) is used:
CC (1) for recombinant production of corresponding proteins; and (2) to
CC produce cells used in screening for modulators. Use of Car and mGluR
CC domains allows presentation of GABABR domains, to a binding agent, in a
CC form more like the natural domain structure compared with use of
CC incomplete receptors, lacking one or more domains. By shuffling different
CC domains, agents can be identified that affect particular domains of a
CC receptor.

SQ Sequence 3234 BP; 739 A; 931 C; 829 G; 735 T; 0 other:

alignment_scores:

Quality: 161.00

Length: 676

Ratio: 0.583 Gaps: 27
Percent Similarity: 40.828 Percent Identity: 20.118
alignment_block:
US-09-775-181-4 x AAZ31049 ..
Align seg 1/1 to: AAZ31049 from: 1 to: 3234
75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
:::|||||:::|||||
394 TCAGAGCACATTCCCTCTACGATTGCTGGTGGGACAACTGGCTCAGG 443
:::|||||:::|||||
91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
:::|||||:::|||||
444 CGTCTCCACGGCAGTGGCAATCTGCTGGGGCTCTCTACATT..... 486
:::|||||:::|||||
108 euProGlyLysTyrProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
|||:::|||||
487CCCCAGGTGAGTTATGCTCTCCCTCCAGC.....AGA 516
|||:::|||||
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG1 141
|||:::|||||
517 CTCCTCAGCAACAAGATCAATTCAAGCTCTTCTCCGGAACC...ATCCC 563
|||:::|||||
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAspTyrG 158
:::|||||:::|||||
564 CAATGATGAGCAGCAGGCCACTGCCATGGCAGACATCATCGAGTATTTC 613
|||:::|||||
158 lnAlaLeuValTrpSerLeuLeuGluGluProSerIleSer..... 172
:::|||||
614 GCTGGAACTGGTGGGCACAAATTGCAGCTGATGACGACTATGGGCGGCCG 663
|||:::|||||
173ArgAlaAlaIleThrPheSerT 180
|||:::|||||
664 GGGATTGAGAAATCCGAGAGGAGCTGAGAAAGGATATCTGCATCGA 713
|||:::|||||
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
|||:::|||||
714 CTTCACTGAACATCATCTCCCACTACTCTGATGAGGAAGAGATCCAGCATG 763
|||:::|||||
196ArgGluGluSer.ArgIleLeuLeuG 204
|||:::|||||
764 TGTAGAGGTGATTCAAAATCCACGGCCCAAGTCACTCGTGGTTTCTCC 813
|||:::|||||
204 lnAspLeuSerSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
|||:::|||||
814 AGTGGCCCATCTTGGCCCTCATCAAGAGATGTCGCGCGCAA... 860
|||:::|||||
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237
:::|||||
861TATCAGCGGCAAGATCTGGCTGGCGCAGGAGCTGGGCCAGCT 904
|||:::|||||
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
|||:::|||||
905 CTCCTCTGATCGCATCTGCTCAGTACTTCCACGTGTTGGCGGCACCAT 954
|||:::|||||
250 rpArgArgLysAspGlyLeuGlyAspLysSerHisPheLysTrpSer 266
|||:::|||||
955 GGATTGCTCTGAAGC.....TGGGCA 977
|||:::|||||
267 ProProTyrLeuGlu.....Cy 272
|||:::|||||
978 GATCCCGAGGCTCCGGGAATTCCTGAAGAAGTCCATCCCGAAGTCTG 1027
|||:::|||||
272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
|||:::|||||
1028 TCCCAATGTTTGGCAGGAGTGTGGGAGAAACATTTAACTGCCAC 1077
|||:::|||||
285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
|||:::|||||
1078 CTCCAAGAAGGTCCAAAGGACCTTTACCTGTGGACACCTTCTTGAGAGG 1127
|||:::|||||
295AsnLeuValProGluPheArg 302

```
1128 TCACGAAGAACTGGCGACAGCTTTAGCAACAGCTCGACAGCTTCCGAC 1177
      |||
302 lyValMetLysValAspIleAsnLeuGlnLysVal
      |||
1178 CCTCTGTACAGGGGATGAGAACATCAGACAGTGTGAGAGCCCTTACATA 1227
      |||
313
      |||
1228 GATTACAGCATTTACGGATATCTTACAAATGTGTACTTAGCAGTCTACTC 1277
      |||
314
      |||
1278 CATTCGCCACGCTTGCAGATATATATACCTGCTTACTCGGAGAGGC 1327
      |||
319
      |||
1328 TCTTCAACAAATGGCTCTGTGCGAGACATCAAGAAAGTTGAGGCGTGGCAG 1377
      |||
325 SerGlyThrHisLysCysHisLeuAsnAsnSer
      |||
1378 GTCCTGAAGCACTACGGCATCTAAACTTTACAAACAATATGGGGAGCA 1427
      |||
336
      |||
1428 GGTGACCTTTGATGAGTGTGGTACCTGGTGGGAACATTCCATCATCA 1477
      |||
348
      |||
1478 ACTGGCACCTCTCCCGAGAGATGCTCCATCGTGTTTAAGGAAGTCGGG 1527
      |||
349 AlaTyrGluCysIleCysLysAlaGly
      |||
1528 TATTACAAAGCTATATGCCAAGAGGAGAAAGACTCTTCAACAGGAGA 1577
      |||
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis
      |||
1578 GAAATCTGTGGAGTGGGTCTCCAGGAGGTGCCCTTCCCAACTGCA 1627
      |||
378
      |||
1628 GCCGAGACTGCTGCGAGGACAGGAAGGATCATTTGAGGGGGAGCCC 1677
      |||
389
      |||
1678 ACCTGCTGTTGAGTGTGTGGAGTGTCTGATGGGAGTATAGTGATGA 1727
      |||
399
      |||
1728 GACAGATGCCAGTGTCTGTACAAAGTCCAGATGACTTCTGGTCCAATG 1777
      |||
406
      |||
1778 AGAACACACCTCTGTCATGCCAAGAGATCGAGTTCTGTCTGGGAGC 1827
      |||
418
      |||
1828 GAGCCCTTTGGGATCGCACTACCTCTTTCGCGTGTGGCATTTTCT 1877
      |||
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
      |||
1878 GACAGCCTTGTGCTGGGTGTGTTTATCAAGTTCGCGAACACACCCATTG 1927
      |||
447 leArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeu 463
      |||
1928 TCAAGCCACCAACAGAGCTCTCTACCTCTCTCTCTCTCTCTCTCTCT 1977
      |||
464 LeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheAr 480
      |||
1978 TGCTCTCTCTCCAGCTCTGTCTCTTCATCGGGGAGCCCCAGGACTGGC 2027
      |||
480 gCysIleLeuLeuArgTTPAlaArgLeuLeuGlyPheAlaThrValTyrG 497
      |||
```

```
2078 GTGCGCGCTGCCCGCCAGCGGCGCTTTGGCATCAGCTTCGTGCTCTCATCT 2077
497 lyThrValThrLeuLysLeuHisArgValLeuLysValPheLeuSerArg 513
      |||
2078 CATGCATCTGCTGTAACCAACCAACCGTGTCTCTGTTGTTGAGGCCAAG 2127
      |||
514 ThrAlaGlnArgIleProTyrMetThrGlyGlyArgValMetArgMetIle 530
      |||
2128 ATCCCCACACAGCTTCCACCGCAAGTGGTGGGGCTCAACCTCAGTTCT 2177
      |||
530 uAlaVal
      |||
2178 GCTGTTTCTCTCTGCACCTTCATGCAGATGTCTATCTGTGTGATCTGGC 2227
      |||
540 heLeuIleGlyThrThrSerSerVal...CysGlnAsnLeuGluLysGln 555
      |||
2228 TCTACACGGCGGCCCTCAAGCTACCGCAACGACGAGCTGGAGGATGAG 2277
      |||
556 IleSerLeuIle
      |||
2278 ATCATCTTTCATCAGCTGCCACGAGGCTCCCTCATGCGCCCTGGGCTTCT 2327
      |||
568 uIlePheAsnMetCysLeuIle 575
      |||
2328 GATCGCTACACCTCCCTGCTG 2349
      |||
seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ51398
seq_documentation_block:
ID_ AAZ51398 standard; cDNA; 3234 BP.
XX
AC AAZ51398;
XX
DT 20-JUN-2000 (first entry)
XX
DE Human wild type calcium sensing receptor, CaSR encoding cDNA.
XX
KW Calcium Sensing Receptor; CaSR; human; calcium level; modulator;
hypotensive; osteopathic; receptor activity; hyperparathyroidism;
KW osteoporosis; Paget's disease; treatment; hypercalcaemia malignancy;
KW hypertension; gene therapy; ss.
XX
OS Homo sapiens.
XX
Key 1..3234
CDS /*tag= a
FT /product= "Human wild type calcium sensing receptor"
FT /note= "Does not include stop codon"
FT /partial
XX
XX W0200006601-A1.
PN
XX
PD 10-FEB-2000.
XX
XX 28-JUL-1999; 99WO-US17116.
XX
XX 30-JUL-1998; 98US-0094702.
XX
XX (AVET ) AVENTIS PHARM PROD INC.
PA
XX Yu KT, Labaudiniere RF, Thrower LW;
PI
XX WPI; 2000-195263/17.
DR
XX P-PSDB; AAY70325.
XX
XX Nucleic acids encoding isoforms of human calcium sensing receptor for
PT treating, e.g. hyperparathyroidism or osteoporosis -
XX
XX Claim 1; Page 70-76; 81pp; English.
XX
XX The present sequence is the cDNA encoding the wild type human calcium
CC sensing receptor CaSR, expressed in the kidney. There are multiple
```


seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV26964

seq_documentation_block: AAV26964 standard; cDNA to mRNA; 3809 BP.

XX AC AAV26964;

XX DT 01-SEP-1998 (first entry)

XX DE Human parathyroid calcium receptor 4.0 gene 4Kb fragment.

XX KW ss; calcium ion concentration; parathyroid hormone; homeostasis;

XX KW kidney; calcium receptor; detection.

XX OS Homo sapiens.

XX Key Location/Qualifiers

XX CDS 373..3809

XX FT US5763569-A.

XX PN 09-JUN-1998.

XX PF 07-JUN-1995; 95US-0484565.

XX PR 07-JUN-1995; 95US-0484565.

XX PR 23-AUG-1991; 91US-0749451.

XX PR 11-FEB-1992; 92US-0834044.

XX PR 21-AUG-1992; 92US-0934161.

XX PR 12-FEB-1993; 93US-0017127.

XX PR 23-FEB-1993; 93US-0009389.

XX PR 22-OCT-1993; 93US-0141248.

XX PR 19-AUG-1994; 94US-0292827.

XX PR 21-OCT-1994; 94WO-US12117.

XX PR 08-DEC-1994; 94US-03353784.

XX (BGHM) BRIGHAM & WOMENS HOSPITAL.

XX PA (NPSP-) NPS PHARM INC.

XX PI Brown EM, Garrett JE, Hebert SC;

XX DR WPI; 1998-347412/30.

XX P-PSDB; AAW54846.

XX Calcium receptor poly:peptide(s) - useful for drug screening or antibody production

XX Example 27; Fig 49; 174pp; English.

XX The Human parathyroid calcium receptor gene encodes a 1078 amino acid protein. The tissue from which the receptor and receptors from bovine parathyroid and rat kidney are derived, respond to changes, and control changes, in calcium ion concentration, e.g. parathyroid hormone regulates Ca2+ homeostasis in blood and extracellular fluid, and kidney function alters through changes in Ca2+ levels in juxtaglomerular and proximal tubule cells in the kidney. The purified receptors (produced recombinantly) can be used to screen for compounds that modulate calcium receptor activity, especially those that can be used to treat diseases associated with the receptors in these tissues. They can also be used to raise antibodies for use in detection assays.

XX Sequence 3809 BP; 910 A; 1071 C; 979 G; 849 T; 0 other;

XX alignment_scores: Quality: 161.00 Length: 676

XX Ratio: 0.583 Gaps: 27

Percent Similarity: 40.828 Percent Identity: 20.118

alignment_block: US-09-775-181-4 x AAV26964

Align seg 1/1 to: AAV26964 from: 1 to: 3809

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91

766 TCAGAGCAGATTCCTCTACGATTGTGTGGGAGCAACATGGCTCAGG 815

91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108

816 CGTCTCCACGGCAGTGGCAATCTGCTGGGCTCTCTACATT 858

108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124

859CCCCAGGTCAATTATGCTCTCCAGC.....AGA 888

125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuGl 141

889 CTCCTCAGCAACAAGATCAATTCAGTCTTCTCCGAACC...ATCCC 935

141 nSerAsnLysSerArgGluGlnInAsnLeuGlnAspLeuAspTyrTrp 158

936 CAATGATGAGCAGCCAGCCATGCCATGCCACATCATCGAGTATTTCC 985

158 InAlaLeuValTrpSerLeuLeuGlyGluProSerIleSer..... 172

986 GCTGGAACCTGGTGGGCACAAATTCAGCTGATGACGACATATGGGGCGG 1035

173ArgAlaAlaIleThrPheSerT 180

1036 GGGATTGAGAAATTCGAGAGGAAGCTGAGGAAAGGGATATCGATCGA 1085

180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195

1086 CTTCACTGAATCATCTCCCACTACTCTGATGAGGAAGAGATCCAGCATG 1135

196ArgGluGluSerArgIleLeuLeu 204

1136 TGGTAGAGGTGATTCAAAATTCACGGCCAAAGTCATCGTGTTCCTCC 1185

204 InAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220

1186 AGTGGCCAGATCTTCAGCCCTCATCAGGAGATTTGCCGGCGCAA... 1232

221 GluTyrPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArg 237

1233TATCAGGGCAAGATCTGGCTGCCAGCGAGGCTGGCCAGCT 1276

237 gGlyProAsnGly.....ProArgGlyLeuGlyHisSerT 250

1277 CCTCCCTGATGCCATTCCTCAGTACTTCCACGTGGTGGCGGCACCAT 1326

250 rParArgLysAspGlyLeuGlyAspLysSerHisPheLysTrpSer 266

1327 GGATTCGCTCTGAAGGC.....TGGGCA 1349

267 ProProTyrLeuGlu.....Cy 272

1350 GATCCACGCTTCGGGAATTCCTGAAGAGTCCATCCAGGAAGTCTG 1399

272 s.GluAsnGlySerTyrLysProGlyTyrTrpLeuValThr..... 284

1400 TCACAAATGGTTTGGCAAGAGATTTTGGGAAGAACAATTAACCTGCCAC 1449

285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294

1450 CTCGAAGAGGTGCAAAAGGACCTTTTACCTGTGGACACCTTTCTGAGAGG 1499

295AsnLeuValProGluPheArg 302

```
1500 TCACGAAGAAGTGGCGACAGGTTTACGACAGCTCGACAGCCCTCCGAC 1549
302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1550 CCCCTCTGTACAGGGGATGAGAACATCAGCAGTGTGAGAGACCCCTTACATA 1599
313 ..... 313
1600 GATTACACGCATTTACGGATATCCTACAATGTGTACTTACAGTCTACTC 1649
314 .....AspIleAspGlnCys..... 318
1650 CATTGCCACCCCTTGCAGATATATATACCTGCTTACCTGGGAGAGGC 1699
319 .....SerSerAspGlyTrpPhe 324
1700 TCTTTACCAATGGCTCTGTGTGCACATCAAGAAAGTTGAGCGGTGCAG 1749
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
1750 GTCCTGAAGCACCTTACGGCATCTAACTTTACAAACAATATGGGGAGCA 1799
336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
1800 GGTGACCTTTGATGAGTGTGTGACCTGGTGGGAACATATTCATCAICA 1849
348 .....Gly 348
1850 ACTGGCACCTCTCCCGACAGATGGCTCCATCGTGTAAAGGAAGTCGG 1899
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProGly 362
1900 TATTACAACTGTATGCAAGAGAGGAGAAAGACTCTTTCATCAACGAGGA 1949
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis..... 377
1950 GAAATCCTGTGGAGTGGTTCACGAGAGGTCCTTCTCCAACATGCA 1999
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
2000 GCCGAGACTGCCTGGCAGGACCAAGGAGGATCATTGAGGGGAGGCC 2049
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
2050 ACCTGCTGTTGAGTGTGTGGAGTGTCTCATGGGAGTATAGTATGA 2099
399 .....CysProPheCysAlaAspAsp..... 405
2100 GACAGATGCCAGTGCCTGTACAGAGTCCCGACAGATGCTTCTGGTCCAATG 2149
406 .....SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
2150 AGAACACACCTCTCTGCATTTGCCAAGGAGATCGAGTTTCTGTGCGGAGC 2199
418 .....LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
2200 GAGCCCTTTGGGATGCGACATCACCTCTTTTGGCGTGTGCTGGCATTTTCT 2249
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
2250 GACAGCCTTTGTGCTGGTGTGTATTACAGTTCGCGAACACACCATTTG 2299
447 leArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeu 463
2300 TCAAGGCCACCAACCGAGACTCTCTACCTCTCTCTCTCTCTCTCTCTCT 2349
464 LeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheAr 480
2350 TGCTGCTTCTCCAGCTCCCTGTCTTCTCATCGGGAGGCCCGAGACTGGAC 2399
480 gCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrG 497
2400 GTCCCGCTTCGCCAGCGCCCTTTGGCATCAGCTTGTGCTGTCTGTCT 2449
```

```
497 lyThrValThrLeuLysLeuHisArgValLeuLysValPheLeuSerArg 513
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2450 CATGCATCTGCTGTAACAAACCAACCGTGTCTCTCTCTCTCTCTCTCTCT 2499
514 ThrAlaGlnArgIleProTyrMetThrGlyGlyArgValMetArgMetLe 530
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2500 ATCCCCACACAGCTTCCACCGCAAGTGTGGGGCTCAACCTGCAGTTCTCT 2549
530 uAlaVal.....IleLeuLeuValValPheTrpP 540
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2550 GCTGCTTTCTCTCTGACCTTCATGACAGATTGTCATCTGTGTGATCTGGC 2599
540 heLeuIleGlyTrpThrSerSerVal...CysGlnAsnLeuGluLysGln 555
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2600 TCTACACCGCGCCCTCAAGCTACCGCAACGAGGAGCTGGAGGATGAG 2649
556 IleSerLeuIle.....GlyGlnGlyLysThrSerAspHisLe 568
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2650 ATCATCTTCATCAGTGCACGAGGGCTCCCTCATGGCCCTGGGCTTCTCT 2699
568 uIlePheAsnMetCysLeuIle 575
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2700 GATCGGCTACCTGCCTGCTG 2721

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAT95859
seq_documentation_block:
ID AAT95859 standard; cDNA to mRNA; 3809 BP.
XX
AC AAT95859;
XX
DT 08-MAY-1998 (first entry)
XX
DE Human parathyroid cell calcium receptor 4.0 (HuPcAr 4.0) cDNA.
XX
KW Human parathyroid cell calcium receptor 4.0; HuPcAr 4.0;
KW calcium homeostasis; hyperparathyroidism; osteoporosis; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 373..3609
FT FT /*tag= a
FT FT /product= parathyroid_cell_calcium_receptor_4.0
XX
PN US5688938-A.
XX
PD 18-NOV-1997.
XX
PF 07-JUN-1995; 95US-0485588.
XX
PR 07-JUN-1995; 95US-0485588.
PR 23-AUG-1991; 91US-0749451.
PR 11-FEB-1992; 92US-0834044.
PR 21-AUG-1992; 92US-0934161.
PR 12-FEB-1993; 93US-0017127.
PR 23-FEB-1993; 93US-0009389.
PR 22-OCT-1993; 93US-0141248.
PR 19-AUG-1994; 94US-0292827.
PR 21-OCT-1994; 94WO-US12117.
PR 08-DEC-1994; 94US-0353784.
XX
PA (BGHM) BRIGHAM & WOMENS HOSPITAL.
PA (NPSF-) NPS PHARM INC.
XX
PI Brown EM, Fuller FH, Garrett JE, Hebert SC;
XX
WPI; 1998-008040/01.
DR P-PSDB; AAW38274.
XX
PT DNA encoding calcium receptor polypeptide(s) - useful for
therapeutic purposes, e.g. hyperparathyroidism and osteoporosis
```

xx Claim 15; Columns 125-134; 174pp; English.
 xx The present sequence encodes human parathyroid cell calcium
 CC receptor 4.0 (hupCar 4.0).
 CC The specification includes details of molecules that can modulate
 CC one or more inorganic ion receptor activities, and antibodies and
 CC antibody fragments targetted to inorganic ion receptor proteins. The
 CC proteins, nucleic acids and antibodies may be used to treat
 CC disorders by modulating one or more inorganic ion receptor
 CC activities, preferably disorders of calcium homeostasis, e.g.
 CC hyperparathyroidism and osteoporosis.
 xx
 SQ Sequence 3809 BP; 910 A; 1071 C; 979 G; 849 T; 0 other;

alignment_scores:
 Quality: 161.00 Length: 676
 Ratio: 0.583 Gaps: 27
 Percent Similarity: 40.828 Percent Identity: 20.118
 alignment_block:
 US-09-775-181-4 x AAT95859 ..
 Align seg 1/1 to: AAT95859 from: 1 to: 3809

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
 766 TCAGAGCATTCCCTCAGATTGCTGTGGTGGAGCACTGGCTCAGG 815
 91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
 816 CGTCTCCACGGCAGTGGCAAAATCTGCTGGGCTCTTCTACATT..... 858
 108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
 859CCCGAGTTCAGTTATGCTCTCTCCAGC.....AGA 888
 125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG 141
 889 CTCCTCAGCAACAGAACTCAATCAAGTCTTTCCCTCCGAACC...ATCCC 935
 141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspAspLeuAspTrpTyrG 158
 936 CAATGATGAGCACCAGCAGCTGCCATGGCAGACATCATCGAGTATTCC 985
 158 InAlaLeuValTrpSerLeuLeuGluGlyGluProSerIleSer..... 172
 986 GCTGGAAGTGGTGGGCACAAATTCGAGCTGATGACGACTATGGGCGCGCG 1035
 173ArgAlaAlaIleThrPheSerT 180
 1036 GGGATTGAGAAATCCGAGAGGAGCTGAGGAAGGATATCTGTCATCGA 1085
 180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
 1086 CTTCACTAGTCACTCTCCAGTACTCTGTATGAGGAGAGATCCAGCATG 1135
 196ArgGluGluSerArgIleLeuLeuG 204
 1136 TGGTAGAGGTGTTCAAAATCCACGGCCAAAGTATCTGTGGTTTCTTCC 1185
 204 InAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
 1186 AGTGCCCCAGATCTTGAGCCCTCATCAAGGAGATTGTCGCGCGCAA... 1232
 221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArg 237
 1233TATCAGGGCAAGATCTGCTGGCCAGCAGGAGCTGGGCGAGCT 1276
 237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
 1277 CCTCCCTGATGCCATGCTCAGTACTTCCACGTTGGTGGCGGCACCAATT 1326

250 rpArgArgLysAspGlyLeuGlyAspLysSerHisPheLysTrpSer 266
 1327 GGATTGCTCTGAGGC.....TGGGCA 1349
 267 ProProTyrLeuGlu.....Cys 272
 1350 GATCCCGAGCTTCGGGAATTCCTCAAGAAGCTCCAGGAAAGTCTG 1399
 272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
 1400 TCCACAATGGTTTTCGCAAGGAGTTTGGGAGAACAATTTAACTGCCAC 1449
 285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
 1450 CTCCAAGAAGGTGCAAAAGGACCTTTACCTGTGGACACCTTTCTGAGAGG 1499
 295AsnLeuValProGluPheArgG 302
 1500 TCACGAAGAAAGTGGCGACAGGTTTAGCAACAGCTCGACAGCCTTCCGAC 1549
 302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
 1550 CCCTCTGTACAGGGATGAGAACATCAGCAGTGTGCGAGACCCCTTACATA 1599
 313 313
 1600 GATTACAGCATTACGGATATCTCAATGTGTACTTAGCAGTCTACTC 1649
 314AspIleAspGlnCys..... 318
 1650 CATTCGCCACGCCCTTCAAGATATATATACCTGTACTTGGGAGAGGGC 1699
 319SerSerAspGlyTrpPhe 324
 1700 TCTTCAACAATGGCTCTCTGTCGAGACATCAACAAGTTGAGCGCTGGCAG 1749
 325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
 1750 GTCTTGAAGCACCCTACGGCATCTAAACTTTACAAACAATATATGGGAGCA 1799
 336GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
 1800 GGTGACCTTTGATGAGTGTGTGTGACCTGTGGGGAACATTCCCATCATCA 1849
 348Gly 348
 1850 ACTGCGACCTCTCCCGCAGAGATGGCTCCATCGTGTTTAAGGAAGTCGGG 1899
 349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProG 362
 1950 TATTACAAGCTATATGCCAAGAAGGAGAAAGACTCTTTCATCAACGAGGA 1949
 362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis.... 377
 1950 GAAATCTCTGTGGAGTGTCTCCAGGAGGTGCCCTTCTCCAACACTGCA 1999
 378IleSerGlySerThrLysAspValSerGluGlu..... 388
 2000 GCCGAGACTGCTGGCAGGACCAGGAAGGATCATTTGAGGGGAGGCC 2049
 389AlaTyrValCysLeuProCysArgGly..... 398
 2050 ACCTGCTGCTTGTAGTGTGTGGAGTGTCTGATGGGAGTATAGTGATGA 2099
 399CysProPheCysAlaAsp..... 405
 2100 GACAGATGCCAGTGTCTTAACAAGTGCCTGAGTACTTCTGGTCCAATG 2149
 406SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
 2150 AGAACCCACACCTCTCTGCATTGCCAAGGAGATCGAGTTTCTGTCGTGGAGG 2199


```
173 .....ArgAlaAlaIleThrPheSerT 180
1036 GGGATTGAGAAATTCGGAGGAGAGCTGAGAAAGGATATCTCGTCGA 1085
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
1086 CTTCAAGTGAACATCTCCAGTACTCTGATCAGGAGAGATCCACCATG 1135
196 .....ArgGluGluSer.ArgIleLeuLeuG 204
1136 TGGTAGAGGTGATTCAAATTCACGGGCAAAAGTCATCGTGGTTTCTCC 1185
204 InAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
1186 AGTGGCCCCAGATCTTGAGCCCCCTCAAGAGGATGTCGCCGCGCAA... 1232
221 GluTrpPheHisGlyLeuArgIlystrpArgProHisLeuHisArgAr 237
1233 .....TATCAGGGCAGATCTGCTGCCAGCGAGGCTGGGCCAGCT 1276
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
1277 CCTCCCTGATGCCATGCTCAGTACTTCCAGTGGTGGCGGCACCAT 1326
250 rpArgArgIlyAspGlyLeuGlyGlyAspLysSerHisPheIlystrpSer 266
1327 GGATTGCTCTGAAGGC.....TGGGCA 1349
267 ProProTyLeuGlu.....Cy 272
1350 GATCCAGGCTCCGGGAATTCCTGAAGAGGTCCATCCAGCAAGATCTG 1399
272 s.GluAsnGlySerTyrlsProGlyTrpLeuValThr..... 284
1400 TCCACAATGGTTTGCAGAGGATTTGGGAAGAAACATTTAACTGCCAC 1449
285 LeuSerSerAlaIleTyrlGlyLeuGlnPro..... 294
1450 CTCGAAGAAGGTGCAAAAGGACCTTTACTGTGGACACCTTTCTGAGAGG 1499
295 .....AsnLeuValProGluPheArgG 302
1500 TCACGAAGAAAGTGGCGAGAGTTTAGCAACAGCTCGACAGCTTCGGAC 1549
302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
1550 CCTCTGTACAGGGGATGAGAACATCAGCAGTGTGAGAGCCCTTACATA 1599
313 ..... 313
1600 GATTACAGCATTTACGGATATCCTACAAATGTGTACTTAGCAGTCTACTC 1649
314 .....AspIleAspGlnCys..... 318
1650 CATTGCCACGCTTCAGAGATATATACCTTGCTTACCTGGGAGAGGCG 1699
319 .....SerSerAspGlyTrpPhe 324
1700 TCTTACCACATGGCTCCTGTGCAGACATCAAGAAAGTTGAGCGGTGGCAG 1749
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
1750 GTCCTGAAGCACCATCAGGCGATCTAAACTTTACAAACAATATATGGGGAGCA 1799
336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
1800 GTTGACCTTTGATGAGTGTGGTGACCTGGTGGGAACATATTCCATCATCA 1849
348 .....Gly 348
1850 ACTGGCACCTCTCCCGAGAGATGGCTCCATCGTGTGTTAAGGAAGTCGG 1899
349 AlaTyrlGluCysIleCysLysAlaGly.....PheTyrlHisProG 362
```

```
1900 TATTACACGCTCTATGCCAAGAGGAGAAACACTCTTTCATCACAGGA 1949
362 yValLeuProValAsnPheArgArgGlyProAspGlnHis.... 377
1950 GAAATCTCTGTGGAGTGGGTTCTCCAGGGAGGTGCCCTTCTCCAACGTCA 1999
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
2000 GCCGAGACTGCCTGGCAGGACAGGAAAGGATCATTTGAGGGGAGGCC 2049
389 .....AlaTyrlValCysLeuProCysArgGluGly..... 398
2050 ACCTGCTGCTTTGAGTGTGGAGTCTCTGATGGGAGATATAGTGATGA 2099
399 .....CysProPheCysAlaAsp..... 405
2100 GACAGATGCCAGTGCCTGTACAAGTGCCAGATGACITCTGTGTCCAATG 2149
406 .....SerProCysPheValGlnGluAspLysTyrlLeuArg..... 417
2150 AGAACACACACCTCTCCATTCGCAAGAGATCGAGTTTCTGCTGGGAGC 2199
418 .....LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
2200 GAGCCCTTTGGGATCGCACTACCTCTTTTGGCTGCTGGGCATTTTCT 2249
430 uAspPheValSerMetLeuValValTyrlHisPheArgLysAlaLysSerI 447
2250 GACAGCCTTTGTGCTGGGTGCTTTTATCAAGTTCGCCACACACCCATTG 2299
447 leArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeu 463
2300 TCAAGGCCACCAACGAGAGCTCTCTACTCTCTCTCTCTCTCTCTCTCT 2349
464 LeuLeuTyrlPheProValIleLeuTyrlPheGluProSerThrPheAr 480
2350 TGCTGCTCTCTCCAGCTCCCTGTTCATPCGGGAGGCCGCCAGGACTGGAC 2399
480 gCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrg 497
2400 GTGCGCGCTGGCGCCAGCGCCTTTGGCATCAGCTTCGTGCTCGCATCT 2449
497 lyThrValThrLeuLysLeuHisArgValLeuLysValPheLeuSerArg 513
2450 CATGCATCCTGTGTGAAACCAACCGTGTCTCTCTGTGTGTGGGCAAG 2499
514 ThrAlaGlnArgIleProTyrlMetThrGlyGlyArgValMetArgMetLe 530
2500 ATCCCCACACAGCTTCCACCGCAAGTGGTGGGGCTCAACCTGCAGTTCT 2549
530 uAlaVal.....IleLeuLeuValValPheTrpP 540
2550 GCTGTTTTCCTCTCGACCTTCATCGAGATTCATCTGTGATCTGGC 2599
540 heLeuIleGlyTrpThrSerSerVal...CysGlnAsnLeuGluLysGln 555
2600 TCTACACGCGCCCCCTCAAGCTACCGCAACACGAGGAGCTGGAGGATGAG 2649
556 IleSerLeuIle.....GlyGlnGlyLysThrSerAspHisLe 568
2650 ATCATCTTTCATCAGCTGCCAGGAGGCTCTCTCATGGCCCTGGGCTTCT 2699
568 uIlePheAsnMetCysLeuIle 575
2700 GATCGGCTACACCTGGCTGCTG 2721
seq_name: /sids1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV82485
seq_documentation_block:
ID: AAV82485 standard; cDNA to mRNA; 3809 BP.
XX
AC AAV82485;
```

XX 19-MAR-1999 (first entry)
 XX Human parathyroid calcium receptor pHPuCAR 4.0 encoding cDNA.
 XX Parathyroid calcium receptor; inorganic ion receptor; osteoporosis;
 XX calcium homeostasis; hyperparathyroidism; seizure; stroke; epilepsy;
 XX spinal cord injury; hypoxia-induced nerve cell damage; cardiac arrest;
 XX neonatal distress; neurodegenerative disease; Alzheimer's disease;
 XX Huntington's disease; Parkinson's disease; dementia; muscle tension;
 XX depression; anxiety; ss.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 XX CDS 373..3609
 XX /*tag= a
 XX
 XX US5858684-A.
 XX 12-JAN-1999.
 XX 07-JUN-1995; 95US-0480751.
 XX 07-JUN-1995; 95US-0480751.
 XX 23-AUG-1991; 91US-0749451.
 XX 11-FEB-1992; 92US-0834044.
 XX 21-AUG-1992; 92US-0934161.
 XX 12-FEB-1993; 93US-0017127.
 XX 23-FEB-1993; 93US-0009389.
 XX 22-OCT-1993; 93US-0141248.
 XX 19-AUG-1994; 94US-0292827.
 XX 21-OCT-1994; 94US-0512117.
 XX 08-DEC-1994; 94US-0353784.
 XX (BGHM) BRIGHAM & WOMEN'S HOSPITAL.
 XX (NPSP-) NPS PHARM INC.
 XX
 XX Balandrin MF, Brown EM, Del Mar EG, Garrett JE;
 XX Hebert SC, Nemeth EP, Van Wagenen BC;
 XX
 XX WPI: 1999-119871/10.
 XX P-PSDB; AAW89565.
 XX
 XX Screening for calcium receptor-active compounds - by recombinant
 XX expression of nucleic acid encoding calcium receptor and determining
 XX the effect of compounds on calcium receptor activity
 XX
 XX Claim 1; Fig 49; 176pp; English.
 XX
 XX A method has been developed of screening for a compound able to affect
 XX one or more activities of a calcium receptor (CR) comprises: (A)
 XX contacting a recombinant cell with a test compound, where the
 XX recombinant cell comprises a recombinant nucleic acid expression the CR,
 XX provided that the cell does not have functional CR expression from
 XX endogenous nucleic acid; (B) determining the ability of the test
 XX compound to affect one or more activities of the calcium receptor; and
 XX (C) comparing the ability with the ability of the test compound to
 XX affect the one or more CR activities in a cell not comprising the
 XX recombinant nucleic acid. The present sequence encodes human
 XX parathyroid CR, designated a pHPuCAR 4.0. The nucleic acid sequence of
 XX pHPuCAR 4.0 can be used as part of the recombinant nucleic acid in the
 XX method described above. The compounds identified can be used to treat
 XX diseases or disorders characterised by abnormal calcium homeostasis, e.g.
 XX hyperparathyroidism, osteoporosis and other bone and mineral-related
 XX disorders. They can also be used for the treatment of diseases and
 XX disorders associated with disrupted Ca²⁺ responses, e.g. seizures,
 XX stroke, spinal cord injury, hypoxia-induced nerve cell damage such as in
 XX cardiac arrest or neonatal distress, epilepsy, neurodegenerative
 XX diseases such as Alzheimer's disease, Huntington's disease and
 XX Parkinson's disease, dementia, muscle tension, depression, and anxiety.
 XX
 XX Sequence 3809 BP; 910 A; 1071 C; 979 G; 849 T; 0 other;

alignment_scores:
 Quality: 161.00 Length: 676
 Ratio: 0.583 Gaps: 27
 Percent Similarity: 40.828 Percent Identity: 20.118
 alignment_block:
 US-09-775-181-4 x AAW82485 ..
 Align seg 1/1 to: AAW82485 from: 1 to: 3809
 75 AlaGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
 766 TCAGAGCACATTCCTCTAGATTGCTGTGGGAGCAACTGGCTCAGG 815
 91 rHisGlnLeuLysArgAlaAsnCyssSerGlyArgTyrGluLeuAlaGlyL 108
 816 CGTCTCCACGGCAGTGGCAATCTGCTGGGGCTCTTCTACATT..... 858
 108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
 859CCCCAGTCAGTTATGCTCTCCAGC.....AGA 888
 125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuL 141
 889, CTCCTCAGCAACAAGAAATCAATTCAGTCTTCTCCTCGAACC...ATCCC 935
 141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAspTrpTyrG 158
 936 CAATGATGAGCACCACGCCACTGCCATGGCAGACATCATCGATGATTTCC 985
 158 In AlaLeuValTrpSerLeuLeuGluGlyGluProSerIleSer..... 172
 986 GCTGGAACTGGTGGGCACATTCGACGCTGATGACGACTATGGCGGCCG 1035
 173ArgAlaAlaIleThrPheSerT 180
 1036 GGGATTGAGAAATTCGAGAGGAGCTGAGGAAAGGATATCTGCATCGA 1085
 180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
 1086 CTTCACTGAACATCTCCAGTACTCTGATGAGGAGATCCAGCATG 1135
 196ArgGluGluSer ArgIleLeuLeuG 204
 1136 TGTAGAGGTGATTCAAATTCACGGCCAAAGTCATCGTGGTTTCTCC 1185
 204 InAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
 1186 AGTGGCCCATCATCTGAGCCCTCATCAGGAGATTGTCGGCGCAA... 1232
 221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237
 1233TATCACCAGGCAAGATCTGGTGGCCAGGAGGCTGGGCCAGCT 1276
 237 ggLYProAsnGlnGly.....ProArgGlyLeuGlyHisSert 250
 1277 CTTCCCTGATCGCATGCTCCCTCAGTACTTCCACGTGGTTGGCGCACCAT 1326
 250 rpArgArgLysAspGlyLeuGlyClyAspLysSerHisPheLysTrpSer 266
 1327 GGATTCGCTGAAGG.....TGGGCA 1349
 267 ProProTyrLeuGlu.....Cy 272
 1350 GATCCAGGCTTCGGGAATTCCTGAAGAGGTCCATCCAGGAAGTCTG 1399
 272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
 1400 TCCACAAATGGTTTGGCAAGGAGTTTGGGAAGAACATTTAACTGCCAC 1449
 285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294


```
|||||..... ||| |||
1450 CTCGAAGAGGTGCAAAAGGACCTTTACCTGTGGACACCTTTCTCGAGAGG 1499
295 ..... AsnLeuValProGluPheArgG 302
1500 TCACGAAGAAAGTGGCAGACAGGTTTAGCAACAGCTCGACAGCCTTCGCAC 1549
302 yValMetLysValAspLeuGlnLysVal..... 313
1550 CCCTCTGTACAGGGGATGAGAACATCATCAGCAGTGTGAGACCCCTTACATA 1599
313 ..... 313
1600 GATTACAGCATTTACGGATATCTTACAAATGTGTACTTAGCAGTCTACTC 1649
314 ..... AspIleAspGlnCys..... 318
1650 CATTCGCCACGCTTGGCAAGATATATATACCTGCTTACCTGGGAGAGGCG 1699
319 ..... SerSerAspGlyTrpPhe 324
1700 TCTTCACCAATGGCTCCTGTGTGCAGACATCAAGAAAGTTGAGGCGTGGCAG 1749
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
1750 GTGCTGAAGCACCTACGGCATCTAAACATTTACAAACAATATGGGGGAGCA 1799
336 ..... GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
1800 GGTGACCTTTGATGAGTGTGTGACCTGGTGGGAACATATTCCATCATCA 1849
348 ..... Gly 348
1850 ACTGCGACCTCTCCCGACAGGATGCTCCATGCTGTTTAAGGAAGTCGGG 1899
349 AlaTyrGluCysIleCysLysAlaGly..... PheTyrHisProG1 362
1900 TATTACACGCTATGCTCAAGAGGGAAGAAAGACTCTTCATCAACGAGGA 1949
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHls.... 377
1950 GAAATCTGTGGAGTGGTCTCCAGGAGGTGGCTTCCTCCAACTGCA 1999
378 ..... IleSerGlySerThrLysAspValSerGluGlu..... 388
2000 GCCGAGACTGCTGCGAGGGACAGAAAGGATCATTTGAGGGGAGGCC 2049
389 ..... AlaTyrValCysLeuProCysArgGluGly..... 398
2050 ACCTGCTGTTTGAAGTGTGTGGAGTGTCTCTGATGGGAGTATAGTGATGA 2099
399 ..... CysProPheCysAlaAspAsp..... 405
2100 GACAGATGCCAGTGCCTGTATCAAGTGGCCAGATGACTTCTGCTCCAATG 2149
406 ..... SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
2150 AGAACACACACCTCTGCTGATGCAAGAGAGATCGAGTTCGTGCTGGAGC 2199
418 ..... LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
2200 GAGCCCTTTGGGATCGCACTACCTCTCTTGGCGTGGCGCATTTTCTCT 2249
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
2250 GACACCCCTTTGCTGGGTGGTGTATATCAAGTTCGCCAACACACCCCATG 2299
447 leArgAlaSerGlyLeuLeLeuLeuGluThrIleLeuPheGlySerLeu 463
2300 TCAAGGCCACCAACGAGAGCTCTCTACCTCTCTCTCTCTCTCTCTCTCT 2349
464 LeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheAr 480
```

```
2350 TGCTGCTTCTCCAGCTCCCTGTTCTTTCATCGGGGAGCCCCAGGACTGGAC 2399
480 gCysIleLeuLeuArgTrpAlaArgLeuGlyPheAlaThrValTyrG 497
2400 GTGCGCGCTGGCGCCAGCGGCTTTGGCATCAGCTTCGTGCTCTCATCT 2449
497 lyThrValThrLeuLysLeuHisArgValLeuLysValPheLeuSerArg 513
2450 CATGCATCTGTTGTAACCAACCAACCGCTGCTCTGCTGTTTGGAGGCCAAG 2499
514 ThrAlaGlnArgIleProTyrMetThrGlyGlyArgValMetArgMetLe 530
2500 ATCCCCACACGCTTCCACCGCAAGTGTGGGGCTCAACCTGCAGTTCT 2549
530 uAlaVal..... IleLeuLeuValValPheTrpP 540
2550 GCTGTTTCTCTGCACCTTTCATGCAGATTGTCATCTGTGTGATCTGCG 2599
540 heLeuIleGlyTrpThrSerSerVal... CysGlnAsnLeuGluLysGln 555
2600 TCTACACCGCGCCCCCTCAAGCTACCGCAACCCAGGAGCTGGAGGATGAG 2649
556 IleSerLeuIle..... GlyGlnGlyLysThrSerAspHisLe 568
2650 ATCATCTTCATCAGCTGCACGAGGCTCCCTCATGCGCCTGGGCTTCT 2699
568 uIlePheAsnMetCysLeuIle 575
2700 GATCGGCTACACCTGCCTGCTG 2721
seq_name: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ89298
seq_documentation_block:
ID AAZ89298 standard; cDNA to mRNA; 3809 BP.
XX
AC AAZ89298;
XX
DT 09-JUN-2000 (first entry)
XX
DE Human calcium receptor phuCaR4.0 cDNA.
XX
KW Calcium receptor; treatment; calcimimetic; calcilytic; osteopathic;
KW cerebroprotective; cytosstatic; neuroprotective; dermatological;
KW tranquilizer; vulnerrary; antituler; immunosuppressive; hypotensive;
KW cardant; parathyroid hormone; osteoporosis; calcitonin secretion;
KW hyperparathyroidism; Paget's disease; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 373..3609
FT FT /*tag= a
FT FT /product= "phuCaR4.0"
XX
PN US6031003-A.
XX
PD 29-FEB-2000.
XX
PF 07-JUN-1995; 95US-0484719.
XX
PR 23-AUG-1991; 91US-0749451.
PR 17-FEB-1992; 92US-0834044.
PR 21-AUG-1992; 92US-0934161.
PR 12-FEB-1993; 93US-0017127.
PR 23-FEB-1993; 93US-0009389.
PR 22-OCT-1993; 93US-0141248.
PR 19-AUG-1994; 94US-0292827.
PR 21-OCT-1994; 94WO-US12117.
PR 08-DEC-1994; 94US-0353784.
XX
PA (NPSP-) NPS PHARM INC.
PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX
```



```
348 ..... Gly 348
1850 ACTGGCAGCTTCCCCAGAGATGCTCCATCGTGTTTAAGAAGTCGGG 1899
349 AlatyrgluCysIleCysIleAlaGly ..... PheTyrHisProG1 362
1900 TATTACAGCTTATGCGCAAGAGGAGAAAGACTCTTTCATCAACGAGGA 1949
362 yValLeuProValAsnAsnPheArgArgArgGlyProAspGlnHis.... 377
1950 GAAATCTCTGTGGAGTGGGTCTCCAGGAGGTGCCCTTCTCCAACACTGCA 1999
378 ..... IleSerGlySerThrLysAspValSerGluGlu..... 388
2000 GCCGAGACTGCTGCCAGGACCAAGAGGATCATTTGAGGGGAGGCC 2049
389 ..... AlaTyrValCysLeuProCysArgGluGly..... 398
2050 ACCTGCTGCTTTGAGTGTGTGGAGTGTCTGTATGGGAGTATAGTGATGA 2099
399 ..... CysProPheCysAlaAsp..... 405
2100 GACAGATGCCAGTGCCTCTAACAGTGCACGATGACTTCTGGTCCAATG 2149
406 ..... SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
2150 AGAACACACTCTCTGCTATCCAGGAGATCGAGATTTCTGTGTGGAGC 2199
418 ..... LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
2200 GAGCCCTTTGGGATCGCACTCACCTCTTTGCCGTGTGGGCATTTTCCT 2249
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
2250 GACAGCCTTTCTGTGTGGTGTGTTTATCAAGTTCGCAACACACCCATTG 2299
447 leArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeu 463
2300 TCAAGGCCACCAACGAGAGCTCTCTACCTCCCTCTCTCTCTCTCTCT 2349
464 LeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheAr 480
2350 TGCTGCTTCTCCAGCTCCTCTCTCTCTCTCGGGAGGCCCCAGGACTGGAC 2399
480 gCysIleLeuLeuArgThrAlaArgLeuLeuGlyPheAlaThrValTyrG 497
2400 CTGCGCGCTGCGCCAGCGCGCTTTGGCATCAGCTTCGTCTGCTCATCT 2449
497 lyThrValThrLeuLysLeuHisArgValLeuLysValPheLeuSerArg 513
2450 CATGATCTCTGGTGAACCAACCGTGTCTCTCTCTCTCTCTCTCTCTCT 2499
514 ThrAlaGlnArgIleProTyrMetThrGlyGlyArgValMetArgMetLe 530
2500 ATCCCCACCACTTCCACCGCAAGTGGTGGGGCTCACTCTGAGTTCTCT 2549
530 uAlaVal..... IleLeuLeuValValPheThrpp 540
2550 GCTGGTTTCTCTCTGCACTTCATGCAGATTTGTCATCTGTGTGATCTGC 2599
540 heLeuIleGlyTrpThrSerVal... CysGlnAsnLeuGluLysGln 555
2600 TCTACCGCGCCGCCCTCAAGCTACCGCAACCCAGGAGCTGGAGGTAG 2649
556 IleSerLeuIle..... GlyGlnGlyLysThrSerAspHisLe 568
2650 ATCATCTTCATCAGTGCACGAGGGCTCTCTCATGGCCCTGGGCTTCTCT 2699
568 uIlePheAsnMetCysLeuIle 575
2700 GATCGCTACACTGCCTGCTG 2721
```

seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ50617

seq_documentation_block:
ID AAZ50617 standard; cDNA; 2703 BP.

```
XX
AC AAZ50617;
XX
DT 20-JUN-2000 (first entry)
XX
DE Human calcium sensing receptor isoform encoding cDNA.
XX
KW Calcium Sensing Receptor; CaSR; isoform; human; splice variant;
KW hypotensive; osteopathic; receptor activity; calcium level; modulator;
KW treatment; hyperparathyroidism; osteoporosis; Paget's disease;
KW hypercalcaemia malignancy; hypertension; gene therapy; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Location/Qualifiers
FT 1..2703
FT /tag= a
FT /product= "Human calcium sensing receptor isoform"
FT /note= "Does not include stop codon"
FT /partial
XX
PN WO200006601-A1.
XX
PD 10-FEB-2000.
XX
PF 28-JUL-1999; 99WO-US17116.
XX
PR 30-JUL-1998; 98US-0094702.
XX
PA (AVET ) AVENTIS PHARM PROD INC.
XX
PI Yu KT, Labaudiniere RF, Thrower LW;
XX
WP1: 2000-195263/17.
DR P-PSDB; AAY45001.
XX
PT Nucleic acids encoding isoforms of human calcium sensing receptor for
treating, e.g. hyperparathyroidism or osteoporosis
XX
PS Claim 5; Page -: 81pp; English.
XX
CC The present sequence is the cDNA encoding an isoform of the human
CC calcium sensing receptor CaSR, expressed in the kidney. It is a splice
CC variant of the wild type CaSR, that arise from deletion of nucleotides
CC 1080-1610, from the extracellular domain, that comprise acidic residues.
CC It has hypotensive and osteopathic activity. The CaSR isoforms can be
CC used to identify agonists and antagonists that modulate the receptor
CC activity and calcium levels. These modulators are useful for treating
CC hyperparathyroidism or osteoporosis, Paget's disease, hypercalcaemia
CC malignancy or hypertension. The DNA sequence is also useful for altering
CC the CaSR activity and in gene therapy.
CC Note: This sequence has deletion of nucleotides 1080-1610, according to
CC the sequence shown in the specification, but has been stated as
CC 1075-1608 in the claims. This sequence is not found in the specification
CC but has been constructed from the human CaSR sequence (Seq ID.No.11)
CC found in page 70-76.
XX
SQ Sequence 2703 BP; 595 A; 804 C; 689 G; 615 T; 0 other;
```

alignment_scores:
Quality: 158.50 Length: 567
Ratio: 0.652 Gaps: 23
Percent Similarity: 42.857 Percent Identity: 20.459alignment_block:
US-09-775-181-4 x AAZ50617

Align seg 1/1 to: AAZ50617 from: 1 to: 2703

75 AlaGluGluValPrometAspValAlaSerTyrLeuTyrThrGlyAspSe 91
:::|||||
394 TCAGAGCACATTCCCTACGATTGCTGGTGGAGCAACTGGCTCAGG 443
91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
:::|||||
444 CGTCTCCAGCGCAGTGCATATCTGCTGGGGCTCTTACATT..... 486
108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
|||:::|||||
487CCCGAGTCAGTTATGCCCTCCAGC.....AGA 516
125 AlaLeuAspThrLeuThrHisAlaThrAspPheLeuAsnValMetLeuG 141
|||:::|||||
517 CTCCTCAGCAACAAGAAATCAATTCAAGTCTTCTCCGGAACC...ATCCC 563
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspAspLeuAspTrpTyrG 158
:::|||||
564 CAATGATGAGCACCAGCCACTGCCATGGGAGAGCATCATCGAGTATTCC 613
158 lnAlaLeuValTrpSerLeuLeuGluGlyGluProSerIleSer..... 172
:::|||||
614 GCTGGAACCTGGTGGGCACAAATTGCCACTGATGACGACTATGGGCGCG 663
173ArgAlaAlaIleThrPheSerT 180
|||:::|||||
664 GGGATTGAGAAATTCCGAGAGGAGCTGAGGAAAGGATATCTGCATCGA 713
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
|||:::|||||
714 CTTCAGTGAACATCTCCCGAGTACTCTGAGGAGAGAGATCCAGCATG 763
196ArgGluGluSerArgIleLeuLeuG 204
|||:::|||||
764 TGGTAGAGGTGATTCAAAATTCACGGCCCAAGTATCGTGGTTCCTCC 813
204 lnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
|||:::|||||
814 AGTGGGCCAGATCTTGAGCCCTCATCAAGAGAGATTGCCGGGCGCAA... 860
221 GluTrpPheHisGlyLeuArgAlaGlyLysTrpArgProHisLeuHisArgAr 237
|||:::|||||
861TATCAGGGGAGAGATCTGGTGGCCAGGCGCTGGGCGCAGCT 904
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
|||:::|||||
905 CCTCCCTGATCGCATGCCCTCAGTACTTCCACGTGGTGGCGCACCAT 954
250 rpArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSer 266
|||:::|||||
955 GGATTCCCTCTGAAGGC.....TGGGCA 977
267 ProProTyrLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeuVa 283
|||:::|||||
978 GATCCAGGCTT.....CCGGGA..... 995
283 lThrLeuSerSerAlaIleTyrGlyLeuGlnProAsnLeuValProGluP 300
|||:::|||||
996ATTCCCTGAAG 1005
300 heArgGlyValMetLysValAspIleAsnLeuGlnLysValAspIleAsp 316
|||:::|||||
1006 AAGGTCCATCCCGAAGATC.....TGTCGA 1031
317 GlnCysSerSerAspGlyTrpPheSerGlyThrHisLysCysHisLeuA 333
|||:::|||||
1032 CAAATGGTTTGGCAAGAGTTTGGGAAGAAACATTTAACTGCCACCTCC 1081
333 snAsnSerGluCysMetProIleLysGlyLeuGlyPheValLeuGlyAla 349
|||:::|||||
1082 CCTTCTCCAACCTGC.....AGC 1098
350 TyrGluCysIleCysLysAlaGlyPheTyrHisProGlyValLeuProVa 366
:::|||||
1099 CGAGACTCC..... 1107
366 lAsnAsnPheArgArgGlyProAspGlnHisIleSerGlySerThrL 383
:::|||||
1108CTGGCAGGGACCAAGGA 1123
383 ysAspValSerGluGlu.....AlaTyrValCysLeuLeuProCys 395
|||:::|||||
1124 AAGGATCAATTGAGGGGAGCCCACTGCTGCTTTGAGTGTGTGGAGTGT 1173
396 ArgGluGly.....CysPropheCy 402
:::|||||
1174 CCTGATGGGAGTATAGTATGATGACAGATGCCAGTGCCTGTATAACAAGTG 1223
402 sAlaAspAsp.....SerProCysPheValGlnG 412
|||:::|||||
1224 CCCAGATGACTTCTGCTGGACGAGGCCCTTTGGGATCGCACTCACCTC 1273
412 luAspLysTyrLeuArg.....LeuAlaIleIleSer 422
|||:::|||||
1274 AGATCGAGTTTCTGCTGGACGAGGCCCTTTGGGATCGCACTCACCTC 1323
423 PheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeuValValTy 439
|||:::|||||
1324 TTTGCCCTGCTGGGCAATTTCCTGACAGCCTTTGCTGGGTGTGTTTAT 1373
439 rHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuIleLeuLeuG 456
|||:::|||||
1374 CAAGTTCGCAACACACACCCATTGTCAAGGCCACCAACCGAGAGCTCCT 1423
456 luThrIleLeuPheGlySerLeuLeuLeuTyrPheProValValIleLeu 472
|||:::|||||
1424 ACCTCCTCCTCTCTCCTCCTGCTGCTGCTGCTTCTCCAGCTCCCTGTTCTTC 1473
473 TyrPheGluProSerThrPheArgCysIleLeuLeuArgTrpAlaArgLe 489
|||:::|||||
1474 ATCGGGGAGCCCGAGGACTGGAGTGGCGCTGGCGCCAGCGGCTTTGG 1523
489 uLeuGlyPheAlaThrValTyrGlyThrValThrLeuLysLeuHisArgV 506
|||:::|||||
1524 CATCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1573
506 alLeuLysValPheLeuSerArgThrAlaGlnArgIleProTyrMetThr 522
|||:::|||||
1574 TCCTCCTGGTGTGGTGGAGCCCAAGATCCCCCAGCTTCCACCGCAAGTGG 1623
523 GlyGlyArgValMetArgMetLeuAlaVal..... 532
|||:::|||||
1624 TGGGGGCTCAACCTGCAGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1673
533 lIleLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerSerVal. 548
|||:::|||||
1674 GATTGTCACTGTGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1723
549 ..CysGlnAsnLeuGluLysGlnIleSerLeuIle.....Gly 560
|||:::|||||
1724 GCACACAGGAGCTGGAGGATGAGATCATCTTCATCAGCTCCACGAGGGG 1773
561 GlnGlyLysThrSerAspHisLeuIlePheAsnMetCysLeuIle 575
:::|||||
1774 TCCCTCATGGCCCTGGGCTTCCCTGATCGGCTACACCTGCCTGCTGCTG 1818

FEATURE:
 NAME/KEY: CDS
 LOCATION: 373..3606
 OTHER INFORMATION:
 US-08-485-588-3

alignment_scores:
 Quality: 161.00 Length: 676
 Ratio: 0.583 Gaps: 27
 Percent Similarity: 40.828 Percent Identity: 20.118
 alignment_block:
 US-09-775-181-4 x US-08-485-588-3 ..
 Align seg 1/1 to: US-08-485-588-3 from: 1 to: 3809

```

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
76 TCAGAGCACATTCCTCTACGATTGCTGGTGGAGCAACTGGCTCAGG 815
91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
816 CGTCTCACGCCAGTGGCAATCTGCTGGGGCTCTTCTACATT..... 858
108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
859 .....CCCCAGTCAGTTATGCTCCTCCAGC.....AGA 888
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuL 141
889 CTCCTCAGCAACAAGATCAATTCAGTCTTTCCTCCGAACC...ATCCC 935
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAspTrpTyrG 158
936 CAATGATGAGCACCAGGCCACTGCCATGGCAGACATCATCGATTATTCC 985
158 lnAlaLeuValTrpSerLeuLeuGlyGluProSerIleSer..... 172
986 GCTGGAACCTGGTGGGCACAATTGCACGTATGATGACGACTATGGCGCGCG 1035
173 .....ArgAlaAlaIleThrPheSerT 180
1036 GGGATTGAGAAATTCAGAGGAAGCTGAGCAAGGATATCTGCATCGA 1085
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
1086 CTTTCAGTGAACATCTCTCCAGTACTCTGATGAGGAGACAGATCCAGCATG 1135
196 .....ArgGluGluSerArgIleLeuLeuG 204
1136 TGGTAGAGGTGATTCAAAATTCACGGCCCAAGTCAATCGTGGTTTCTCC 1185
204 lnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
1186 AGTGGCCCATGATCTTGAGCCCTCAATCAAGGAGATTGCTCCGGCGCAA... 1232
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237
1233 .....TATCAGGGGCAAGATCTGGCTGGCCAGCGAGGCTTGGCCAGCT 1276
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
1277 CCTCCCTGATCGCCATGCCCTCAGTACTTCCACGCTGGTGGCGGCACCAT 1326
250 rArgArgLysAspGlyLeuGlyClyAspLysSerHisPheLysTrpSer 266
1327 GGAATCCGCTGAAGGC.....TGGGCA 1349
267 ProProTyrLeuGlu.....Cy 272
1350 GATCCCAAGGCTTCGGGAATTCCTGAAGAAGGTCCATCCCAAGAACTCTG 1399

```

```

272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
1400 TCCACAATGGTTTGGCCAGGAGTTTGGGAAGAACAATTAACTGCCAC 1449
285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
1450 CTCACAAGAGGTGCAAAAGACCTTTTACCTGTGGACACCTTTCTGAGAGG 1499
295 .....AsnLeuValProGluPheArgG 302
1500 TCAGGAAGAAGTGGCGACAGAGTTTACCAACAGCTCGACAGCCTTCCGAC 1549
302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
1550 CCCTCTGTACAGGGGATGAGAACAATCAGCAGTGTGAGAGACCCCTTACATA 1599
313 ..... 313
1600 GATTACACGCATTACGGATATCCTACAATGTGTACTTAGCAGTCTACTC 1649
314 .....AspIleAspGlnCys..... 318
1650 CATTCGCCACGCCTTGCAGATATATATACCTGCTTACCTGGGAGAGGGC 1699
319 .....SerSerAspGlyTrpPhe 324
1700 TCTTACCAATGGCTCCTGTGCAGACATCAAGAAAGTTGAGGCGTGCAG 1749
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
1750 GTCTGTAAGCACCTCAGCGCATCTAAACTTTACAACAATATGGGGGAGCA 1799
336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
1800 GGTGACCTTTGATGAGTGTGCTGACCTGGTGGGAACTATTCATCATCA 1849
348 .....Gly 348
1850 ACTGGCACCTCTCCCGACAGAGATGGCTCCATCGTGTTTAAGGAAGTCGGG 1899
349 AlaTyrGluCysLecysLysAlaGly.....PheTyrHisProG 362
1900 TATTACAACGCTATGCCAAGAAGGGAGAAAGACTCTTTCATCAACAGGAG 1949
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis.... 377
1950 GAAATCCTGTGGAGTGGGTCTCCAGGGAGGTGCCCTTCTCCAAGTGA 1999
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
2000 GCCGAGACTGCCTGGCAGGACCAGGAAGGATCATTTGAGGGGAGGCC 2049
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
2050 ACCTGCTGCTTGTGAGTGTGGAGTCTCCTGATGGGAGTATAGTGATGA 2099
399 .....CysProPheCysAlaAspAsp..... 405
2100 GACAGATGCCAGTGCCTGTAAACAAGTCCCGAGATGACTTCTGTGTCNAATG 2149
406 .....SerProCysPheValGlnGlnAspLysTyrLeuArg..... 417
2150 AGAACACACCTCCTGCATTGCCAAGGAGATCGAGTTTCTGCTGGGAGC 2199
418 .....LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
2200 GAGCCCTTTGGGATCGCACTCACCCCTCTTTTGGCTGCTGGGCATTTCTCT 2249
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
2250 GACAGCCTTTGTGCTGGTGTGTTTATCAAGTTCCGCAACACACCATTTG 2299
447 leArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeu 463

```



```
313 ..... 313
1600 GATTACACGATTACGGATATCCTACAATGTGTACTTAGCAGTCTACTC 1649
314 .....AspIleaspGlnCys..... 318
1650 CATTGCCACCGCCTTGCAGATATATATACCTGCTTACCTGGGAGAGGC 1699
319 .....SerSerAspGlyTrpPhe 324
1700 TCTTACCAATGCTCTGTCGACACATCAAGAAAGTTGAGGCGTGCAG 1749
325 SerGlyThrHisLysCysHisLeuAsnSer..... 335
1750 GTCCTGAAGCACCTACCGCATCTAAACTTTACAAACAATATGGGGAGCA 1799
336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
1800 GGTGACCTTTGATGAGTGTGGTACCTGGTGGGGAACATTATCCATCATCA 1849
348 .....Gly 348
1850 ACTGGCACCTCTCCCGACAGGATGGCTCCATCGTGTTTAAGGAAGTCGGG 1899
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProG1 362
1900 TATTACAACGCTCTATGCCAAGAAGGGAGAAAGACTCTTTCATCAACGAGGA 1949
362 yValLeuProValAsnPheArgArgGlyProAspGlnHis..... 377
1950 GAAATCCTGTGGAGTGGTTCGAGGGAGTGGCCCTTCCCACTGCA 1999
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
2000 GCCGAGACTCCCTGGCAGGACCAAGGAAGGATCATTTGAGGGGAGCCCC 2049
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
2050 ACCTGTGCTTGTAGTGTGTGGAGTGTCTGTATGGGAGTATAGTGATGA 2099
399 .....CysProPheCysAlaAsp..... 405
2100 GACAGATGCCAGTCCCTGTGAACAGTCCCGCAGATGACTTCTGTGTCCAATG 2149
406 .....SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
2150 AGAACACACCTCTCTGATTCGCAAGGAGATCGAGTTTCTGTCTGGTGGACG 2199
418 .....LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
2200 GAGCCCTTTGGGATCGCAGCTCACCCCTTTTGGCGTGTGGGCAATTTCTCT 2249
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
2250 GACAGCCTTTGTCTGGTGTGTATCAAGTTCGCCAAGACACACCCATTCG 2299
447 leArgAlaSerGlyLeuIleLeuGluThrIleLeuPheGlySerLeu 463
2300 TCAAGGCCACCAACCGAGAGCTCTCTACCTCTCTCTCTCTCTCTCTCTCT 2349
464 LeuLeuTyrPheProValIleLeuTyrPheGluProSerThrPheAr 480
2350 TGCTGTCTCTCCAGCTCCCTGTTCTTCTCATCGGGAGCCCGAGGACTGGAC 2399
480 gCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrG 497
2400 GTCCCGCCTTCGCCGCGCCCTTTGGCATCAGCTTCGTCTGTCTGTCTGT 2449
497 lyThrValThrLysLeuHisArgValLeuLysValPheLeuSerArg 513
2450 CATGCATCTGTGTGAACCAACCGTGTCTCTGTGTGTGTGTGTGTGTGTGTGT 2499
```

```
514 ThrAlaGlnArgIleProTyrMetThrGlyGlyArgValMetArgMetLe 530
2500 ATCCCCACACAGCTTCCACCGCAAGTGTGGGGCTCAACCTGCAGTTCTCT 2549
530 uAlaVal.....IleLeuLeuValValPheTrpP 540
2550 GCTGGTTTCTCTCTGACCTTCATGAGATGTCATCTGTGTGATCTGGC 2599
540 heLeuIleGlyTrpThrSerSerVal...CysGlnAsnLeuGluLysGln 555
2600 TCTACACCGCGCCCTCAAGCTACCGCAACAGGAGCTGGAGGATGAG 2649
556 IleSerLeuIle.....GlyGlnGlyLysThrSerAspHisLe 568
2650 ATCATCTTCATCAGTCCACGAGGCTCCTCATGGCCCTGGGCTTCTCT 2699
568 uIlePheAsnMetCysLeuIle 575
2700 GATCGGCTACACCTGCTGCTG 2721

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-943-986-3

seq_documentation_block:
; Sequence 3, Application US/08943986
; Patent No. 5962314
; GENERAL INFORMATION:
; APPLICANT: Edward M. Brown
; APPLICANT: Steven C. Hebert
; APPLICANT: James E. Garrett, Jr.
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; CITY: 633 West Fifth Street
; STATE: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTSEQ
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,986
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/484,565
; FILING DATE: 7-June-1995
; APPLICATION NUMBER: 08/353,784
; FILING DATE: 9 December, 1994
; APPLICATION NUMBER: PCT/US/94/12117
; FILING DATE: 21 October, 1994
; APPLICATION NUMBER: U.S. 08/292,827
; FILING DATE: 23 August, 1994
; APPLICATION NUMBER: U.S. 08/141,248
; FILING DATE: 22 October, 1993
; APPLICATION NUMBER: U.S. 08/009,389
; FILING DATE: 23 February, 1993
; APPLICATION NUMBER: U.S. 08/017,127
; FILING DATE: 12 February, 1993
; APPLICATION NUMBER: U.S. 07/934,161
; FILING DATE: 21 August, 1992
; APPLICATION NUMBER: U.S. 07/834,044
; FILING DATE: 11 February, 1992
; APPLICATION NUMBER: U.S. 07/749,451
; FILING DATE: 23 August, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Heber, Sheldon O.
```

```

; REGISTRATION NUMBER: 38,179
; REFERENCE/DOCKET NUMBER: 213/006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 373..3606
; OTHER INFORMATION:
;
; US-08-943-986-3

```

```

alignment_scores:
  Quality: 161.00      Length: 676
  Ratio: 0.583         Gaps: 27
Percent Similarity: 40.828 Percent Identity: 20.118

```

```
alignment_block:

```

```
US-09-775-181-4 x US-08-943-986-3 ..
```

```
Align seg 1/1 to: US-08-943-986-3 from: 1 to: 3809
```

```

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
   :::::::::::::: :::::
76 TCAGAGCACATTCCCTTACGATTGCTGTGGGAGCAACTGGCTCAGG 815
   :::::::::::::: :::::
91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
   :::::::::::::: :::::
816 CGTCTCCACGGCAGTGGCAAACTCTGCTGGGCTCTCTACATT..... 858
   :::::::::::::: :::::
108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
   :::::::::::::: :::::
859 .....CCCCAGGTCAGTTATGCTCTCCAGC.....AGA 888
   :::::::::::::: :::::
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuGl 141
   :::::::::::::: :::::
889 CTCCTCAGCAACAAGAAATCAATTCAAGTCTTCTCCGAACC...ATCCC 935
   :::::::::::::: :::::
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspAspLeuAspTrpTyrG 158
   :::::::::::::: :::::
936 CAATGATGAGCACCAGGCCACTGCCATGCCAGACATCATCGAGTATTCC 985
   :::::::::::::: :::::
158 InAlaLeuValTrpSerLeuLeuGluGlyGluProSerIleSer..... 172
   :::::::::::::: :::::
986 GCTGGAACCTGGGTGGGCACAAATTGACGCTGATGACGACTATGGCGCGCG 1035
   :::::::::::::: :::::
173 .....ArgAlaAlaIleThrPheSerT 180
   :::::::::::::: :::::
1036 GGGATTGAGAAATTCGAGAGGAAGCTGAGGAAGGGATATCTGCATCGA 1085
   :::::::::::::: :::::
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
   :::::::::::::: :::::
1086 CUTCAGTCAACTATCTCCCACTACTCTGATGATGAGGAAGAGATCCAGATG 1135
   :::::::::::::: :::::
196 .....ArgGluGluSerArgIleLeuLeuG 204
   :::::::::::::: :::::
1136 TGGTAGAGTGATTCAAAATTCACGGCCAAAGTCATCGTGGTTTCTCC 1185
   :::::::::::::: :::::
204 InAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
   :::::::::::::: :::::
1186 AGTGGCCAGATCTTGAGGCCCTCATCAGGAGATTGTCCCGCGCAA... 1232
   :::::::::::::: :::::
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237
   :::::::::::::: :::::
1233 .....TATCAGGGCAAGATCTGGCTGGCCAGGAGGCCCTGGGCCAGCT 1276

```

```

237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
   :::::::::::::: :::::
1277 CCTCCTGATGCCCATGCTCAGTACTTCCACGTGGTGGCGGCACCAATT 1326
   :::::::::::::: :::::
250 rpArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSer 266
   :::::::::::::: :::::
1327 GGATTGCTCTCAAGGC.....TGGGCA 1349
   :::::::::::::: :::::
267 ProProTyrLeuGlu.....Cy 272
   :::::::::::::: :::::
1350 GATCCAGGCTTCGGGAATTCCTCAAGAGGTCCATCCAGGAAGTCTG 1399
   :::::::::::::: :::::
272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
   :::::::::::::: :::::
1400 TCACAAATGGTTTGGCAAGGAGTTTGGGAAGAAACATTTAACTGCCAC 1449
   :::::::::::::: :::::
285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
   :::::::::::::: :::::
1450 CTCCAAGAGGTGCAAAAGGACCTTTACCTGTGGACACCTTTCTGAGAGG 1499
   :::::::::::::: :::::
295 .....AsnLeuValProGluPheArg 302
   :::::::::::::: :::::
1500 TCACGAAGAAAGTGGCGACAGGTTTAGCAACAGCTCGACAGCCTTCCGAC 1549
   :::::::::::::: :::::
302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
   :::::::::::::: :::::
1550 CCTCTGTACAGGGGATGAGAACATCAGCAGTGTGAGAGCCCTTACATA 1599
   :::::::::::::: :::::
313 ..... 313
1600 GATTACACGATTTACGGATATCTTACAATGTGTACTTAGCAGTCTACTC 1649
   :::::::::::::: :::::
314 .....AspIleAspGlnCys..... 318
   :::::::::::::: :::::
1650 CATTGCCACGCTTGCAGATATATATACCTGCTTACCTGGGAGAGGCG 1699
   :::::::::::::: :::::
319 .....SerSerAspGlyTrpPhe 324
   :::::::::::::: :::::
1700 TCTTCAACCAATGGCTCTGTGCAGACATCAAGAAAGTTGAGCGGTGGCAG 1749
   :::::::::::::: :::::
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
   :::::::::::::: :::::
1750 GTCCTGAAGCACCTACGGCATCTAACTTTACAAACAATATATGGGAGCA 1799
   :::::::::::::: :::::
336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
   :::::::::::::: :::::
1800 GGTGACCTTTGATGAGTGTGTGACCTGTGGGGAACATATTCATCATCA 1849
   :::::::::::::: :::::
348 .....Gly 348
   :::::::::::::: :::::
1850 ACTGGCACCTCTCCCGAGAGGATGGCTCCATCGTGTTTAAGGAAGTCGGG 1899
   :::::::::::::: :::::
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProGl 362
   :::::::::::::: :::::
1900 TATTACAAGCTCTATCCCAAGAGGAGAAAGACTCTTTCATCAACGAGGA 1949
   :::::::::::::: :::::
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis.... 377
   :::::::::::::: :::::
1950 GAAATCTGTGGAGTGGTTCCTCAGGAGGAGTGCCCTTCTCCAACATGCA 1999
   :::::::::::::: :::::
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
   :::::::::::::: :::::
2000 GCCGAGACTGCTCGCAGGACCAAGAGGATCATATTGAGGGGGAGCCC 2049
   :::::::::::::: :::::
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
   :::::::::::::: :::::
2050 ACCTGCTGCTTGTAGTGTGTGGAGTCTCTGATGGGAGTATAGTATGA 2099
   :::::::::::::: :::::
399 .....CysProPheCysAlaAsp..... 405
   :::::::::::::: :::::
2100 GACAGATGCCAGTGGCTGTAAAGTGGCCAGATGACTTCTGGTCCAATG 2149

```

```
406 .....SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
2150 AGAACACACCTCTCTCCAGTCCGACGAGATCGATTCTCTGCTGACG 2199
418 .....LeuAlaIleSerPheGlnGlyLeuCysMetLeuLe 430
2200 GAGCCCTTTGGGATCGCAGCTCTTTGCCGCTGCTGGCATTTCCT 2249
430 uAspPheValSerMetLeuValTyrHisPheArgLysAlaLysSerI 447
2250 GACAGCCCTTTGCTGGTGTGTATATCAAGTTCGCAACACACCCATTG 2299
447 leArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeu 463
2300 TCAAGGCCACACACGAGAGCTCTCCACCTCTCTCTCTCTCTCTCTC 2349
464 LeuLeuTyrPheProValIleLeuTyrPheGluProSerThrPheAr 480
2350 TGCTGCTCTCCAGCTCCCTGCTCTCTCATCGGGGAGCCCGAGGCTGAC 2399
480 gCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrG 497
2400 GTGCCGCTCGCCGAGCGGCTTTGGCATCAGCTTCTGCTGCTGCTGCT 2449
497 lyThrValThrLeuLysLeuHisArgValLeuLysValPheLeuSerArg 513
2450 CATGCATCTCTGGTGAACCAACCGCTCTCTCTCTCTCTCTCTCTCT 2499
514 ThrAlaGlnArgIleProTyrMetThrGlyGlyArgValMetArgMetLe 530
2500 ATCCCCACGAGTCCACCCGAGTGTGGGGCTCACTGAGTTCCT 2549
530 uAlaVal.....IleLeuLeuValValPheThrpp 540
2550 GCTGCTTTCTCTGACCTTCATGCAGATTGTCATCTGTGTCATCTGCG 2599
540 heLeuIleGlyTrpThrSerSerVal...CysGlnAsnLeuGluLysGln 555
2600 TCTACACCGCGCCCTCAAGTACCGCAACAGGAGCTGGAGGATGAG 2649
556 IleSerLeuIle.....GlyGlnGlyLysThrSerAspHisLe 568
2650 ATCATCTTCATCAGTCCGAGGGGCTCTCATGCGCCCTGGGCTTCCT 2699
568 uIlePheAsnMetCysLeuIle 575
2700 GATCGGCTACACCTGCTGCTG 2721
```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:us-08-353-784-3

seq_documentation_block:

Sequence 3, Application US/08353784

Patent No. 6011068

GENERAL INFORMATION:

APPLICANT: Edward F. Nemeth, Edward M.

APPLICANT: Brown, Steven C. Hebert,

APPLICANT: Bradford C. Van Wagenen, Manuel

APPLICANT: F. Balandrin, Forrest H. Fuller,

APPLICANT: Eric G. DelMar, and Scott T. Moe

TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE

MOLECULES

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon

STREET: First Interstate World Center

STREET: Suite 4700

STREET: 633 West Fifth Street

CITY: Los Angeles

STATE: California

COUNTRY: USA

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,784
FILING DATE: 9 December, 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 8
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 209/069
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3809 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 373..3606
OTHER INFORMATION:
US-08-353-784-3
```

alignment_scores:

Quality: 161.00

Ratio: 0.583

Percent Similarity: 40.828

Percent Identity: 20.118

alignment_block:

US-09-775-181-4 x US-08-353-784-3

Align seg 1/1 to: US-08-353-784-3 from: 1 to: 3809

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspse 91

766 TCAGGACACATTCCTCTACGATTGTGTGGGAGCACTGGCTCAGG 815

91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108

816 CGTCTCCAGCGCAGTGCATAATCTGCTGGGCTCTTCTACATT..... 858

108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124

859CCCCAGGTGAGTTATGCTCTCCCTCCAGC.....AGA 888

125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG 141

889 CTCCTCAGCAACAAGATCAATCAAGTCTTCTCCCGAACC....ATCCC 935


```

|||||      |||||:::
2700 GATCGGCTACACTGCCTGCTG 2721

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-484-719B-3

seq_documentation_block:
; Sequence 3., Application US/08484719B
; Patent No. 6031003
; GENERAL INFORMATION:
; APPLICANT: Edward F. Nemeth, Edward M.
; APPLICANT: Brown, Steven C. Hebert,
; APPLICANT: Bradford C. Van Wagonen,
; APPLICANT: Manuel F. Baladrin,
; APPLICANT: Forrest H. Fuller, Eric G.
; APPLICANT: Delmar, Scott T. Moe
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; MOLECULES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS Word
; SOFTWARE: FASTSEQ for Windows Version 3.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,719B
; FILING DATE: 7 June, 1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/353,784
; FILING DATE: 9 December, 1994
; APPLICATION NUMBER: PCT/US/94/12117
; FILING DATE: 21 October, 1994
; APPLICATION NUMBER: U.S. 08/292,827
; FILING DATE: 23 August, 1994
; APPLICATION NUMBER: U.S. 08/141,248
; FILING DATE: 22 October, 1993
; APPLICATION NUMBER: U.S. 08/009,389
; FILING DATE: 23 February, 1993
; APPLICATION NUMBER: U.S. 08/017,127
; FILING DATE: 12 February, 1993
; APPLICATION NUMBER: U.S. 07/934,161
; FILING DATE: 21 August, 1992
; APPLICATION NUMBER: U.S. 07/834,044
; FILING DATE: 11 February, 1992
; APPLICATION NUMBER: U.S. 07/749,451
; FILING DATE: 23 August, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Douglas C. Murdock
; REGISTRATION NUMBER: 37,549
; REFERENCE/DOCKET NUMBER: 213/007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 373..3606

```


; TELEPHONE: (213) 489-1600
 ; TELEFAX: (213) 955-0440
 ; TELE: 67-3510
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3809 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA to mRNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 373...3606
 ; OTHER INFORMATION:
 ; US-08-546-998-2

alignment_scores:
 Quality: 161.00 Length: 676
 Ratio: 0.583 Gaps: 27
 Percent Similarity: 40.828 Percent Identity: 20.118
 alignment_block:
 US-09-775-181-4 x US-08-546-998-2 ..
 Align seg 1/1 to: US-08-546-998-2 from: 1 to: 3809

```

75 AlaGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 TCAGAGCAGATCCCTACGATTGCTGGTGGGAGCAACTGCTCAGG 815
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
91 rHisGlnLeuLysArgAlaAsnCysserGlyArgTyrGluLeuAlaGlyL 108
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
816 CGTCTCCAGCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 858
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
859 .....CCCCAGGTGAGTTATGCTCCTCCAGC.....AGA 888
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG1 141
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
889 CTCCTCAGCAACAAGATCAATTCAGTCTTCTCCGAACC...ATCCC 935
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAspTrpTyrG 158
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
936 CAATGATGAGCAGCAGCCACTGCCATGGCAGACATCATCGAGTATTTC 985
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
158 lnAlaLeuValTrpSerLeuLeuGluGlyGluProSerIleSer..... 172
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
986 GCTGGAAGTGGTGGGCAATTCGAGCTGATGACGACTATGGGCGCGC 1035
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
173 .....ArgAlaAlaIleThrPheSerT 180
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1036 GGGATTGAGAAATCCGAGAGGAGCTGAGGAAGGATATCTGCATCGA 1085
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1086 CTTTCAGTGAACATCTCCAGTACTCTGATGAGGAAGAGATCCAGCATG 1135
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 .....ArgGluGluSer.ArgIleLeuLeuG 204
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1136 TGTTAGAGGTGATTCAAAATTCACCGCCCAAGTCATCGTGGTTCCTCC 1185
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 lnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1186 AGTGGCCAGATCTTGAGCCCTCATCAGGAGATTGTCGGCGCAA... 1232
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1233 .....TATCAGCGGCAAGATCTGGCTGGCGCAGCGAGGCTGGCGCAGCT 1276
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

1277 CCTCCCTGATCGCCATCGCTCAGTACTTCCACGTTGGTGGCGGACCACTT 1326
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
250 rArgArgLysAspGlyLeuGlyAspLysSerHisPheLysTrpSer 266
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1327 GGATTCCTCTGAAGGC.....TGGGCA 1349
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
267 ProProTyrLeuGlu.....Cy 272
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1350 GATCCAGGCTTCGGGAATTCCTGAAGAAGTCCATCCAGGAAGTCTG 1399
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1400 TCCACAATGTTTGGCCAGGAGTTTGGGAAGAAACATTTAACTGCCAC 1449
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1450 CTCCAAGAAGTGC AAAAGGACCTTTACCTGTGGACACCTTTCTGAGAGG 1499
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
295 .....AsnLeuValProGluPheArgG 302
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1500 TCAGGAAGAAGTGGCGACAGGTTTAGCAACAGCTCGACAGCCTTCGGAC 1549
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1550 CCCTCTGTACAGGGGATGAGAACATCAGCAGTGTGGAGACCCCTTACATA 1599
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
313 ..... 313
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1600 GATTACAGCATTTACGGATATCTCAATCTGTACTTACAGTCTACTC 1649
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 .....AspIleAspGlnCys..... 318
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1650 CATTTGCCACGCCCTTGCAAGATATATATACCTTACCTTGGGAGAGGC 1699
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
319 .....SerSerAspGlyTrpPhe 324
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1700 TCTTCAACATGGCTCTCTGTGCAGACATCAAGAAAGTTGAGCGTGGCAG 1749
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1750 GTCTGAGCACCCTACGGCATCTAAACTTTACAAACAATATGGGGAGCA 1799
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
336 .....GluCysMetProLysGlyLeuGlyPheValLeu. 347
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1800 GGTGACCTTTGATGAGTGTGTGACCTGTGGTGGGAACTATTCCATCATCA 1849
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
348 .....Gly 348
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1850 ACTGGCACCTCTCCCGAGAGGATGGTCCATCGTGTAAAGGAAGTCGGG 1899
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProG1 362
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1900 TATTACAACGCTCTATGCCAAGAAGGAGAAAGACTCTTCATCAACGAGGA 1949
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis.... 377
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1950 GAAATCCTGTGAGTGGTGTCTCCAGGGAGGTGCCCTTCTCCAACGTCA 1999
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2000 GCGGAGACTGCCTGGCAGGAGCAGGAAAGGATCATTTGAGGGGAGGCC 2049
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2050 ACCTGCTGCTTTGAGTGTGTGAGTGTCTCTGAGGGAGTATAGTATGA 2099
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
399 .....CysProPheCysAlaAsp..... 405
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2100 GACAGATGCCAGTGCCTGTAAACAAGTCCAGATGACTTCTGTGTCCAATG 2149
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
406 .....SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2150 AGAACACACCTCTCTCATTTGCCAAGGAGATCGAGTTTCTGTGCGGACG 2199
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



```

|||||
1148 ACAATTGAGCTGACGATGACTATGGCGGCCAGGAGTCCGAGAAGTTTCG 1197
83 laSerTyrLeuTyrThrGlyAspSerHisGlnLeuLysArgAlaAsnCy 99
|||||
1198 AGAGGAGCTGAGGAGGGACATCTGCATCGACTTCAGCGAGCTCAT.. 1245
99 sSerGlyArgTyrGluLeuAlaGlyLeuPro.....GlyLysT 112
|||||
1246 .....CTCCCAATACTCTGATGAGGAAAAG 1270
112 rpProAlaLeuAlaSerAlaHisProSerLeuHisArgAlaLeuAspThr 128
|||||
1271 ATCCAGCAGGTGGTGGAGGTGATCCAGAAATCCACCGCAA..... 1311
129 LeuThrHisAlaThrAsnPheLeuAsnValMetLeuGlnSerAsnLysSe 145
|||||
1312 ...AGTCATTTGCT.....CTTCAGCGGCCAGAC 1343
145 rArgGluGlnAsnLeuGlnAspAspLeuAspTrpTyrGlnAlaLeuVal 161
|||||
1344 TGGAAACCCCTCATCAAGAGATCTCCGGCGCAATATCACAGGCGAGATC 1393
162 TrpSerLeuLeuGluGlyGluProSerIleSerArgAlaAla..... 175
|||||
1394 TGGCTGGCCAGCGAGGCTGGCGAGCTCTCCCTGATTGTCATGCCCGA 1443
176 .....IleThrPheSerThrAspSerLeuS 184
|||||
1444 GTATTTCCATGTGTCGGAGCCACCATGGTTGGTTGAAAGCTGGGC 1493
184 erAlaProAlaProGlnValPheLeuGlnAlaThrArgGluGluSerArg 200
|||||
1494 AGATCCAGGCTCCCGGAATCTCTCGAGAAAGTCCACCCAGCAAGTCT 1543
201 IleLeuLeuGlnAspLeuSerSerAlaProIleHisLeuAlaAsnAlaTh 217
|||||
1544 GTC.....CACAATGGTTTGGCCAGGAGTTTGGGAAGAAAC 1581
217 rLeuGluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisL 234
|||||
1582 ATTTAACTGCCACCTGCAAGAGGTGTAAAGCCCATTTACCGGTGGACA 1631
234 euHisArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrp 250
|||||
1632 CTTCTCTGAGAGTCCAGAAAGAGAGGTGCCAGGTTTAAGCAACAGTCCC 1681
251 ArgArgLysAspGlyLeu...GlyGlyAspLysSerHisPheLysTrpSe 266
|||||
1682 ACTGCCTTCCGACCTCTGTGCACCTGGGGAGGAGAACATCAGCAGTGTGCA 1731
266 rProTyrTrpLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeuV 283
|||||
1732 GACTCTTACATGGATTATACACATTTACGGATATCTACACAGCTTACT 1781
283 alThrLeuSerSerAlaIleTyrGlyLeuGlnPro.....AsnLeuVal 297
|||||
1782 TAGCGGTCTACTCCATTGCTCATGCCCTACAAGATATATACACCTGCATA 1831
298 ProGluPheArgGlyValMetLysValAspIleAsnLeuGlnLysValAs 314
|||||
1832 CCTGG...AGAGGGCTCTTC.....ACCAACGGTTCTCTGCGCAGA 1869
314 pIleAspClnCysSerSerAspGlyTrpPheSerGlyThrHisLysCysH 331
|||||
1870 TATCAAGAAGTT.....GAAGCTTGGCAGGTCTCGTAACACACCTGGGC 1913
331 isLeuAsnAsnSer.....GluCys 337
|||||
1914 ACCTAAATTTTACCAGCAATATGGGGAGCAAGTAACATTTTCGATGAATGT 1963
338 MetProIleLysGlyLeuGlyPheValLeu..... 347
|||||

```

```

1964 GGAGACCTGGCAGGAACTATTCCATCATCAACTGGCACCTCTCCCCAGA 2013
348 .....GlyAlaTyrGluCysIleCysL 355
|||||
2014 GGACGGCTCCTAGTGTTTAAGGAAGTTGGATATACAATGTCTATGCCA 2063
355 ysAlaGly.....PheTyrHisProGlyValLeuProValAsnAsn 368
|||||
2064 AGAAGCAGAGAGACTCTTTCATCAATGATGAAAAAATCTGTGGAGTGA 2113
369 PheArgArgGlyProAspGlnHis.....IleSerG 380
|||||
2114 TTCTCAAGGAGGTGCTTTCTCAACTGCAGTGCAGACTGCCTGGCAGG 2163
380 ySerThrLysAspValSerGluGlu.....AlaTyrValCysL 393
|||||
2164 GACCAGAAAGGAATCAATTGAGGGGAGCCCACTGCTGCTTTGAGTGTG 2213
393 euProCysArgGluGly.....Cys 399
|||||
2214 TGGAAATGCTGTGATGGGAGGTACAGCGACGACAGATGCAAGTGCCTGT 2263
400 ProPheCysAlaAspAsp.....SerProCysPh 409
|||||
2264 GATAAGTCCCTGTGATGACTTCTGTGTCATGAGAACCAACACACTTCTGCAT 2313
409 eValGlnGluAspLysTyrLeuArg.....LeuAlaI 420
|||||
2314 CGCAAGCAGATCGAGTTTCTGTGGGACCGAGCCCTTCGGGATCGCAC 2363
420 leIleSerPheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeu 436
|||||
2364 TCACGCTCTTTGTGCTGGGATTTTCTCACAGCCTTCGTCGTGGGC 2413
437 ValValTyrHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuI 453
|||||
2414 GTTCTCATCAAGTTCGCAACAGCGCCATCGTCAAGGCCACCAACCGGGA 2463
453 eLeuLeuGluThrIleLeuPheGlySerLeuLeuTyrPheProValV 470
|||||
2464 CTTCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2513
470 alIleLeuTyrPheGluProSerThrPheArgCysIleLeuLeuArgTrp 486
|||||
2514 TGTTCCTTCATCGGGAGCCCGACGACGTGGCGCTCCGCGCCAGCCG 2563
487 AlaArgLeuLeuGlyPheAlaThrValTyrGlyThrValThrLeuLysLe 503
|||||
2564 GCCTTTGGCATCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2613
503 uHisArgValLeuLysValPheLeuSerArgThrAlaGlnArgIleProT 520
|||||
2614 CAATCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2663
520 yrMetThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLeuVal 536
|||||
2664 GGAAGTGTGGGGCTCAACCTCGACTTCCTGCTGCTCTCTCTCTCTCTCT 2713
537 ValPheTrpPheLeuIleGly.....TrpThrSerSerValCysGlnAs 551
|||||
2714 TTCATCAGATTGTCATCTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2763
551 nLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAspHisL 568
|||||
2764 GACCTACCGCAACCAACGAGCTG.....GAGACGACGAGA 2795
568 euIlePheAsnMetCys 573
|||||
2796 TCATCTTTCATCACCTGC 2812

```

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-484-565-1

seq_documentation_block:

Sequence 1, Application US/08484565
Patent No. 5763569
GENERAL INFORMATION:
APPLICANT: Edward M. Brown
APPLICANT: Steven C. Hebert
APPLICANT: James E. Garrett, Jr.
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,565
FILING DATE: 7 June, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 9
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/006
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5275 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 515..3769
OTHER INFORMATION:
US-08-484-565-1

alignment_scores: Quality: 152.50 Length: 657
Ratio: 0.524 Gaps: 27

Percent Similarity: 44.292 Percent Identity: 21.005
alignment_block:
US-09-775-181-4 x US-08-484-565-1
Align seg 1/1 to: US-08-484-565-1 from: 1 to: 5275
24 AlaSerArgAspProGln.....GlyArgProAspSe 34
||| ::::| ||| |||
954 GCTGGGCATCTCCACAGCAGTGGCCAACTGCTGGGGCTCTTACATC 1003
34 rProArgGluArgThrProLysGlyLysProHisAlaGlnProGlyA 51
|||||::: ||| |||
1004 CCCAGGTACGTATGCT.....CTCCAGCAGACTCCTCAGCAACA 1047
51 rGAlaSerAlaSerAspSerSerAlaProTyr..... 61
|| ::::| ||| |||
1048 GAATCAATCAAGTCTCTCCCGCACCATACCCCAATGATGAACACCA 1097
62SerArgSerThrAspGlyThrIleLeuAl 71
||| ::::| ||| |||
1098 CCACGGCCATGGCTGACATCATCGAGTACTTCGCTGGAAGTGGTGGC 1147
71 aGln.....LysLeuAlaGluValProMetAspValA 83
||| ::::| ||| |||
1148 ACAATTGCAGTGAGGATGATGGCGGCGGATCGAGAAGTTTCG 1197
83 lAsrTyrLeuTyrThrGlyAspSerHisGlnLeuLysArgAlaAsnCy 99
||| ||| ::::| ||| |||
1198 AGAGGAAGCTGAGGAGGACATCTGCATCGACTTCAGCGAGCTCAT.. 1245
99 sSerGlyArgTyrGluLeuAlaGlyLeuPro.....GlyLysT 112
1246CTCCCAATCTCTGATGAGGAAAG 1270
112 rProAlaLeuAlaSerAlaHisProSerLeuHisArgAlaLeuAspThr 128
||||| ::::| ||| |||
1271 ATCCAGCAGGTGGTGGAGGTATCCAGAAATCCACCCGCAA..... 1311
129 LeuThrHisAlaThrAsnPhelAsnValMetLeuGlnSerAsnLysSe 145
:::| ::::| ||| |||
1312 ...AGTCATTGCTG.....CTTCAGCGGCCAGACC 1343
145 rArgGluGlnAsnLeuGlnAspAspLeuAspTrpTyrGlnAlaLeuVal 161
:::| ::::| ||| |||
1344 TGGAAACCCCTCATCAAGAGATCGTCGCGCAATATCACAGGAGGATC 1393
162 TrpSerLeuGluGlyGluProSerIleSerArgAlaAla..... 175
||| ||| ::::| ||| |||
1394 TGGTGGCCAGCGAGCTGGCGCAGCTCTCCCTGATTGCTATGCCGA 1443
176IleThrPheSerThrAspSerLeu 184
||| ||| ::::| ||| |||
1444 GTATTTCCATGTTGCGGAGGCGCACCATTGGTTGGTTTGAAGCTGGC 1493
184 erAlaProAlaProGlnValPheLeuGlnAlaThrArgGluGluSerArg 200
:::| ::::| ||| |||
1494 AGATCCAGGCTTCGGGAATTCCTGCAGAAATCCACCCAGGAAGTCT 1543
201 lLeuLeuGlnAspLeuSerSerSerAlaProHisLeuAlaAsnAlaTh 217
:::| ::::| ||| |||
1544 GTC.....CACAAATGGTTTGCAGAGGAGTTTGGGAAGAACA 1581
217 rLeuGluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisL 234
||| ::::| ||| |||
1582 ATTTAACTGCCACCTGCAGAGGCTGCTAAAGGCCCATACCGGTGGACA 1631
234 euHisArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrp 250
||||| ::::| ||| |||
1632 CCTTCTCAGAGGTACCAAGAAGGAGGTGCAGGTTAAGCAACAGCTCC 1681
251 ArgArgLysAspGlyLeu...GlyGlyAspLysSerHisPheLysTrpSe 266
||| ||| ::::| ||| |||

1682 ACTGCCCTTCCACCTCTGTGCATCGGGAGGAGAACATCAGCAGTGTGCA 1731
266 rProPtyrLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeuV 283
1732 GACTCCTTACATGATATACATTTACGATATCTCACAACGTCTACT 1781
283 alThrLeuSerAlaIleTyrClyLeuGlnPro.....AsnLeuVal 297
1782 TAGCGGTCTACTCCTATGCTCATGCCCTACAAGATATATACACCTGCATA 1831
298 ProGluPheArgGlyValMetLysValAspIleAsnLeuGlnLysValas 314
1832 CTTGGG...AGAGGCTCTTC.....ACCAACGGTTCTCCGCGCAGA 1869
314 pIleAspGlnCysSerSerAspGlyTrpPheSerGlyThrHisLysCysH 331
1870 TATCAAGAAGTT.....GAAGCTTGGCAGGCTCTGAAACACCTGCGGC 1913
331 IseLeuAsnAsnSer.....GluCys 337
1914 ACCTAATTTTACGACGAATATGGGGAGCAAGTAAGCTTTTCGATGAATGT 1963
338 MetProIleLysGlyLeuGlyPheValLeu..... 347
1964 GGACACTGCCAGGAACATACTTCATCATCACTGCACCTCTCCCCAGA 2013
348GlyAlaTyrGluCysIleCysL 355
2014 GGACGGTCCATAGTGTTTAAGGAAGTTGGATATACAAATGCTATGCCA 2063
355 ysAlaGly.....PheTyrHisProGlyValLeuProValAsnAsn 368
2064 AGAAGAGAGAGACTCTTTCATCAATGATGAAAAAATTTCTGTGGAGTGA 2113
369 PheArgArgGlyProAspGlnHis.....IleSerG1 380
2114 TTCTCAAGGGAGGTGCTTCTCAACTGCAGTCGACACTGCGCTGGCAGG 2163
380 ySerThrLysAspValSerGluGlu.....AlaTyrValCysL 393
2164 GACCAGGAAGGAATCATTTAGGGGGNGCCCACTGCTGTTGAGTGTG 2213
393 euProCysArgGlyGly.....Cys 399
2214 TGAATGTCTGTATGGGGATACAGCGAGACAGATGCAAGTGCCTGT 2263
400 ProPheCysAlaAsp.....SerProCysPh 409
2264 GATAAGTGCCTGATGACTTCTGTGTCATGAGAACACACTTCTTCGCAT 2313
409 eValGlnGluAspLysTyrLeuArg.....LeuAlaI 420
2314 CGCCAAGGAGATCGAGTTTCTGTGTGGACCGAGCCCTTCGGGATCGCAC 2363
420 leIleSerPheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeu 436
2364 TCACGCTCTTGTGCTGCTGGCATTTCTCACAGCCTTCTGCTGCTGGGC 2413
437 ValValTyrHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuI 453
2414 GTTTCATCAAGTTCGCGAACACGCCCATCGTCAAGGCCACCAACCGGA 2463
453 eLeuLeuGluThrIleLeuPheGlySerLeuLeuTyrPheProValV 470
2464 GCTCTCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2513
470 alIleLeuTyrPheGluProSerThrPheArgCysIleLeuLeuArgTrp 486
2514 TGTTCCTTCGCGGAGCCCGAGGACTGGACGTGCGCTGCGCCAGCG 2563
487 AlaArgLeuLeuGlyPheAlaThrValTyrGlyThrValThrLeuLysLe 503
2564 GCTTTGGCATCAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2613

503 uHisArgValLeuLysValPheLeuSerArgThrAlaGlnArgIleProt 520
2614 CAATCGGTCCTCTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2663
520 yrMetThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLeuVal 536
2664 GGAAGTGGTGGGCTCAACCTGCAGTTCCTGCTGCTGCTGCTGCTGCTG 2713
537 ValPheTrpPheLeuIleGly.....TrpThrSerSerValCysGlnAs 551
2714 TTTCATGAGATTCTCATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2763
551 nLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAspHisL 568
2764 GAGCTACCGCAACACGAGCTG.....GAGGACGAGA 2795
568 eullePheAsnMetCys 573
2796 TCATCTTCATCACCTGC 2812
seq_name: /cgn2_6ptodata/2/lna/5B_COMB.seq:US-08-480-751-1
seq_documentation_block:
; Sequence 1, Application US/08480751
; Patent No. 5858684
; GENERAL INFORMATION:
; APPLICANT: Edward F. Nemeth
; APPLICANT: Edward M. Brown
; APPLICANT: Steven C. Hebert
; APPLICANT: Forrest H. Fuller
; APPLICANT: James E. Garrett, Jr.
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; STREET: First Interstate World Center
; STREET: Suite 4700
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTSEQ
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,751
; FILING DATE: 7 June, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 9
; APPLICATION NUMBER: 08/353,784
; FILING DATE: 9 December, 1994
; APPLICATION NUMBER: PCT/US/94/12117
; FILING DATE: 21 October, 1994
; APPLICATION NUMBER: U.S. 08/292,827
; FILING DATE: 23 August, 1994
; APPLICATION NUMBER: U.S. 08/141,248
; FILING DATE: 22 October, 1993
; APPLICATION NUMBER: U.S. 08/009,389
; FILING DATE: 23 February, 1993
; APPLICATION NUMBER: U.S. 08/017,127
; FILING DATE: 12 February, 1993
; APPLICATION NUMBER: U.S. 07/934,161
; FILING DATE: 21 August, 1992
; APPLICATION NUMBER: U.S. 07/834,044
; FILING DATE: 11 February, 1992
; APPLICATION NUMBER: U.S. 07/749,451

seq_documentation_block:
 ; Sequence 1, Application US/08353784
 ; Patent No. 6011068
 ; GENERAL INFORMATION:
 ; APPLICANT: Edward F. Nemeth, Edward M.
 ; APPLICANT: Brown, Steven C. Hebert, Manuel
 ; APPLICANT: Bradford C. Van Wagenen, Manuel
 ; APPLICANT: F. Balandrin, Forrest H. Fuller,
 ; APPLICANT: Eric G. DelMar, and Scott T. Moe
 ; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
 ; TITLE OF INVENTION: MOLECULES
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Lyon & Lyon
 ; STREET: First Interstate World Center
 ; STREET: Suite 4700
 ; STREET: 633 West Fifth Street
 ; CITY: Los Angeles
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 90071
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: FASTSEQ
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/353,784
 ; FILING DATE: 9 December, 1994
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA: including application
 ; PRIOR APPLICATION DATA: described below: 8
 ; APPLICATION NUMBER: PCT/US/94/12117
 ; FILING DATE: 21 October, 1994
 ; APPLICATION NUMBER: U.S. 08/292,827
 ; FILING DATE: 23 August, 1994
 ; APPLICATION NUMBER: U.S. 08/141,248
 ; FILING DATE: 22 October, 1993
 ; APPLICATION NUMBER: U.S. 08/009,389
 ; FILING DATE: 23 February, 1993
 ; APPLICATION NUMBER: U.S. 08/017,127
 ; FILING DATE: 12 February, 1993
 ; APPLICATION NUMBER: U.S. 07/934,161
 ; FILING DATE: 21 August, 1992
 ; APPLICATION NUMBER: U.S. 07/834,044
 ; FILING DATE: 11 February, 1992
 ; APPLICATION NUMBER: U.S. 07/749,451
 ; FILING DATE: 23 August, 1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Heber, Sheldon O.
 ; REGISTRATION NUMBER: 38,179
 ; REFERENCE/DOCKET NUMBER: 209/069
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (213) 489-1600
 ; TELEFAX: (213) 955-0440
 ; TELEX: 67-3510
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5275 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA to mRNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 515..3769
 ; OTHER INFORMATION:
 ; US-08-353-784-1

alignment_scores: Quality: 152.50 Length: 657

Ratio: 0.524 Gaps: 27
 Percent similarity: 44.292 Percent identity: 21.005
 alignment_block:
 US-09-775-181-4 x US-08-353-784-1 ..
 Align seg 1/1 to: US-08-353-784-1 from: 1 to: 5275
 24 AlaSerArgAspProGln.....GlyArgProAspSe 34
 ||| :|||
 954 GCTCGGCATCTCCACAGCAGTGGCCCAACCTGCTGGGGCTTCTTACATC 1003
 34 rProArgGluThrProLysGlyLysProHisAlaGlnGlnProGlyA 51
 |||||:|
 1004 CCCAGGTACGTATGCT.....CTCCAGCAGACTCTCTCAGCAACA 1047
 51 rGAlaSerAlaSerAspSerAlaProTrp..... 61
 |||:|
 1048 GAATCAATCAAGTCCTTCTCCGACCATACCCCAATGATGACACACAGG 1097
 62SerArgSerThrAspGlyThrIleLeuAl 71
 |||:|
 1098 CCACGGCCATGGCTGACATCATCGACTTCCGCTGGAACCTGGTGGC 1147
 71 aGln.....LysLeuAlaGluGluValProMetAspValA 83
 |||:|
 1148 ACAATTGCAGCTGACGATGACTATGGCCGCCAGGATCGAGAAGTTTCG 1197
 83 laSerTyrLeuTyrThrGlyAspSer.HisGlnLeuLysArgAlaAsnCy 99
 |||||:|
 1198 AGAGGAAGCTGAGGAGGAGCATCTGCATCGACTTCAGCGAGTCAAT.. 1245
 99 sSerGlyArgTyrGluLeuAlaGlyLeuPro.....GlyLysT 112
 |||||
 1246CTCCCAATACTCTGATGAGGAAAG 1270
 112 rpProAlaLeuAlaSerAlaHisProSerLeuHisArgAlaLeuAspThr 128
 |||||:|
 1271 ATCCAGCAGGTGCTGGAGGTGATCCAGAAATTCACGCCCAA..... 1311
 129 LeuThrHisAlaThrAsnPheLeuAsnValMetLeuGlnSerAsnLysSe 145
 :|||:|
 1312 ...AGTCATGTCTG.....CTTCTCCAGCGGCCAGACC 1343
 145 rArgGluGlnAsnLeuGlnAspAspLeuAspTrpTyrGlnAlaLeuVal 161
 :|||:|
 1344 TGGACCCCTCATCAAGAGATCGTCCGGCGCAATATCACAGCAGCATC 1393
 162 TrpSerLeuLeuGluGlyGluProSerIleSerArgAlaAla..... 175
 |||:|
 1394 TGGCTGCCAGCGAGGCTGGCCAGCTCTCCCTGATGCTATGCCCGA 1443
 176IleThrPheSerThrAspSerLeuS 184
 |||:|
 1444 GTATTTCCATGTGTCGGAGGCACCATTTGGTTTGTGAAAGCTGGGC 1493
 184 verAlaProAlaProGlnValPheLeuGlnAlaThrArgGluGluSerArg 200
 :|||:|
 1494 AGATCCAGGCTTCCGGGAATCTCTCCAGAAAGTCCACCCAGGAAGTCT 1543
 201 IleLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaTh 217
 :|||:|
 1544 GTC.....CACAAATGTTTTCAGAGGAGTTTGGGAAGAAAC 1581
 217 rLeuGluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisL 234
 |||||:|
 1582 ATTTAACTGCCACCTGCAAGAGGTTGCTAAAGGCCCATTTACCGGTGGACA 1631
 234 euHisArgArgGlyProAsnGlnGlnArgGlyLeuGlyHisSerTrp 250
 |||||:|
 1632 CCTCTCTGAGAGGTACAGAGAGGAGGTCCAGAGTTAAGCAACAGTCCC 1681
 251 ArgArgLysAspGlyLeu...GlyGlyAspLysSerHisPheLysTrpSe 266


```
1682 ACTGCTTCCGACCTCTGTGCACTGGGAGGAGAACATCAGCAGTGTGCA 1731
      ||| |||.....:
266  rProProTyrLeuGluCysGluAsnGlySerTyrLysProGlyTyrLeuV 283
      : |||.....:
1732 GACTCCTTACATGGATTATACACATTTACGGATATCCACACAGTCTACT 1781
      : |||.....:
283  alThrLeuSerSerAlaIleTyrGlyLeuGlnPro.....AsnLeuVal 297
      ||| |||.....:
1782 TAGCGCTCTACTCCATTCGTATGCTGCCTCAACAGATATATACACCTGCATA 1831
      ||| |||.....:
298  ProGluPheArgGlyValMetLysValAspIleAsnLeuGlnLysValAs 314
      ||| |||.....:
1832 CTGGG...AGAGGCTCTTC.....ACCAACGGTTCCTGCGCAGA 1869
      ||| |||.....:
314  pIleAspGlnCysSerSerAspGlyTyrPheSerGlyThrHisLysCysH 331
      ||| |||.....:
1870 TATCAAGAAGTT.....GAAGCTTGGCAGGTCCTGAAACACCTGCGGC 1913
      ||| |||.....:
331  IsLeuAsnAsnSer.....GluCys 337
      ||| |||.....:
1914 ACCTAAATTTTACCAGCAATATATGGGAGCAAGTAACCTTTCGATGAATGT 1963
      ||| |||.....:
338  MetProIleLysGlyLeuGlyPheValLeu..... 347
      ||| |||.....:
1964 GGAGACCTGGCAGGGAAGTATTTCATCATCACTGGCACCTCTCCCCAGA 2013
      ||| |||.....:
348  .....GlyAlaTyrGluCysIleCysL 355
      ||| |||.....:
2014 GCACGGCTCCATAGTGTTAAGCAAGTTGGATATTACAAATGCTATGCCA 2063
      ||| |||.....:
355  ysAlaGly.....PheTyrHisProGlyValLeuProValAsnAsn 368
      ||| |||.....:
2064 AGAAAGAGAGAGACTCTTCATCAATGATGAAAAAATCTGTGGAGTGA 2113
      ||| |||.....:
369  pheArgArgGlyProAspGlnHis.....IleSerGln 380
      ||| |||.....:
2114 TTCTCAAGGGAGGTGCTTTCFCAACTGCATCGAGACTGCTGGCAGG 2163
      ||| |||.....:
380  ySerThrLysAspValSerGluGlu.....AlaTyrValCysL 393
      ||| |||.....:
2164 GACCAGGAAGAAATCATTTAGGGGGAGCCCACTGCTGCTTGTAGTGTG 2213
      ||| |||.....:
393  euProCysArgGluGly.....Cys 399
      ||| |||.....:
2214 TGAATGCTCTGATGGGAGTACAGGAGACAGATGCAAGTGCCTGT 2263
      ||| |||.....:
400  ProPheCysAlaAsp.....SerProCysPh 409
      ||| |||.....:
2264 GATAAGTCCCTGATGACTTCTGTGTCAGCCAGCCCTTCGGGATCGCAC 2313
      ||| |||.....:
409  eValGlnGluAspLysTyrLeuArg.....LeuAlaI 420
      ||| |||.....:
2314 CCGCAAGAGATCGAGTTCTGCTGTGACCCAGCCCTTCGGGATCGCAC 2363
      ||| |||.....:
420  leIleSerPheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeu 436
      ||| |||.....:
2364 TCACGCTCTTCTGCTGTGGCATTTCCTCACAGCCTTCGTGTGGGC 2413
      ||| |||.....:
437  valValTyrHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuIl 453
      ||| |||.....:
2414 GCTTTCATCAAGTTCGCAACAGCCCACTGCTCAAGGCCACCAACCGGA 2463
      ||| |||.....:
453  eLeuLeuGluThrIleLeuPheGlySerLeuLeuLeuTyrPheProValV 470
      ||| |||.....:
2464 GCTCTCCTATCTCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2513
      ||| |||.....:
470  alIleLeuTyrPheGluProSerThrPheArgCysIleLeuLeuArgTrp 486
      ||| |||.....:
2514 TGTCTTCTATCGGGAGCCCGGAGGACTGGACGTGCGGCTGGCCAGCGG 2563
      ||| |||.....:
487  AlaArgLeuLeuGlyPheAlaThrValTyrGlyThrValThrLysLe 503
      ||| |||.....:
2564 GCCTTTGGCATCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2613
      ||| |||.....:
503  uHisArgValLeuLysValPheLeuSerArgThrAlaGlnArgIleProT 520
      : |||.....:
2614 CAATCGGTCCTCTCTGTTTGGAGCCCAAGATTCCACACAGCTTCCACC 2663
      : |||.....:
520  yzMetThrGlyGlyArgValMetArgMetLeuAlaValIleLeuVal 536
      ||| |||.....:
2664 GGAAGTGGTGGGGCTCAACCTGCATGCTGCTGCTGCTGCTGCTGCTG 2713
      ||| |||.....:
537  ValPheTyrPheLeuIleGly.....TrpThrSerSerValCysGlnAs 551
      ||| |||.....:
2714 TTCATGCAGATTGTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2763
      ||| |||.....:
551  nLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAspHisL 568
      : |||.....:
2764 GAGTACCGCAACACGAGCTG.....GAGGACGAGA 2795
      : |||.....:
568  euIlePheAsnMetCys 573
      : |||.....:
2796 TCATCTTTCATCACCTGC 2812
      : |||.....:
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-484-719B-1
seq_documentation_block:
: Sequence 1, Application US/08484719B
: Patent No. 6031003
: GENERAL INFORMATION:
: APPLICANT: Edward F. Nemeth, Edward M.
: APPLICANT: Brown, Steven C. Hebert,
: APPLICANT: Bradford C. Van Wagenen,
: APPLICANT: Manuel F. Balandrin,
: APPLICANT: Forrest H. Fuller, Eric G.
: APPLICANT: Delmar, Scott T. Moe
: TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
: TITLE OF INVENTION: MOLECULES
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lyon & Lyon
: STREET: First Interstate World Center
: STREET: Suite 4700
: STREET: 633 West Fifth Street
: CITY: Los Angeles
: STATE: California
: COUNTRY: USA
: ZIP: 90071
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: MS Word
: SOFTWARE: FastSeq for Windows Version 3.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/484,719B
: FILING DATE: 7 June, 1995
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/353,784
: FILING DATE: 9 December, 1994
: APPLICATION NUMBER: PCT/US/94/12117
: FILING DATE: 21 October, 1994
: APPLICATION NUMBER: U.S. 08/292,827
: FILING DATE: 23 August, 1994
: APPLICATION NUMBER: U.S. 08/141,248
: FILING DATE: 22 October, 1993
: APPLICATION NUMBER: U.S. 08/009,389
: FILING DATE: 23 February, 1993
: APPLICATION NUMBER: U.S. 08/017,127
: FILING DATE: 12 February, 1993
: APPLICATION NUMBER: U.S. 07/934,161
: FILING DATE: 21 August, 1992
: APPLICATION NUMBER: U.S. 07/834,044
: FILING DATE: 11 February, 1992
: APPLICATION NUMBER: U.S. 07/749,451
```

: FILING DATE: 23 August, 1991
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Douglas C. Murdock
 : REGISTRATION NUMBER: 37, 549
 : REFERENCE/DOCKET NUMBER: 213/007
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (213) 489-1600
 : TELEFAX: (213) 955-0440
 :

alignment_scores:		
Quality:	152.50	Length: 657
Ratio:	0.524	Gaps: 27
Percent Similarity:	44.292	Percent Identity: 21.005

```

alignment_block:
  US-09-775-181-4 x US-08-484-719B-1 ..
  Align seg 1/1 to: US-08-484-719B-1 from: 1 to: 5275

24 AlaSerArgAspProGln.....GlyArgProAspse 34
   ||| ::||| |||
   954 GCTCGGGCATCTCCACAGCAGTGGCAACCTGCTGGGCTCTTCTACATC 1003

34 rProArgGluArgThrProLysGlyLysProHisAlaGlnProGlyA 51
   |||||::: ||| ::::: |
   1004 CCCCAGGTTCAGTATGCGCT.....CTCCAGCAGACTCCTCGACCAAA 1047

51 rgAlaSerAlaSerAspSerSerAlaProTrp..... 61
   || ::::: ||| ||||| |||||:::
   1048 GAATCAATTCAGTCTCTCTCCGACCACTACCCAATGATGAACACGAGG 1097

```

[illegible]

seq_name: /cqn2_6/ptodata/2/ina/5A_COMB.seq:US-08-485-588-2

```

alignment_scores:
    Quality: 150.00      Length: 690
    Ratio: 0.543        Gaps: 27
    Percent Similarity: 40.000    Percent Identity: 19.565

alignment_block:
US-08-775-181-4 x US-08-485-588-2 ..

Align seg 1/1 to: US-08-485-588-2 from: 1 to: 5006

      75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
829 TCAGAGCACATTCCCTCTACGATTGCTGTGGTGGGAGCAACTGGCTCAGG 878

      91 rHisGlnLeuLysArgAlaAsnCysSerGluLysArgTyrGluLeuAlaGlyL 108
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
879 CGTCTCACGGCAGTGGCAAACTGCTGGGGCTCTCTTACATT..... 921

      108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
922 .....CCCCAGTCAGTTATGCGCTCTCTCCAGC.....AGA 951

      125 AlaLeuAspThrLeuThrHisAlaThrAsnProPheLeuAsnValMetLeuGl 141
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
952 CTCCTCAGCAACAAGAAATCAATTCAAGTCTTTCTCTCCGAACC...ATCC 998

```


2751 CATCAGTCCACGAGGGCTCCCTCATGCGCTGGGCTTCCTGATCGGCT 2800

571 snMetCysLeuIle 575

|||||:

2801 ACACCTGCGTCTG 2814

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq.us-08-484-565-2

seq_documentation_block:

; Sequence 2, Application US/08484565

; Patent No. 5763569

; GENERAL INFORMATION:

; APPLICANT: Edward M. Brown

; APPLICANT: Steven C. Hebert

; APPLICANT: James E. Garrett, Jr.

; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE

; MOLECULES

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: First Interstate World Center

; STREET: Suite 4700

; STREET: 633 West Fifth Street

; CITY: Los Angeles

; STATE: California

; COUNTRY: USA

; ZIP: 90071

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: FASTSEQ

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/484,565

; FILING DATE: 7 June, 1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; PRIOR APPLICATION DATA: including application

; PRIOR APPLICATION DATA: described below: 9

; APPLICATION NUMBER: 08/353,784

; FILING DATE: 9 December, 1994

; APPLICATION NUMBER: PCT/US/94/12117

; FILING DATE: 21 October, 1994

; APPLICATION NUMBER: U.S. 08/292,827

; FILING DATE: 23 August, 1994

; APPLICATION NUMBER: U.S. 08/141,248

; FILING DATE: 22 October, 1993

; APPLICATION NUMBER: U.S. 08/009,389

; FILING DATE: 23 February, 1993

; APPLICATION NUMBER: U.S. 08/017,127

; FILING DATE: 12 February, 1993

; APPLICATION NUMBER: U.S. 07/934,161

; FILING DATE: 21 August, 1992

; APPLICATION NUMBER: U.S. 07/834,044

; FILING DATE: 11 February, 1992

; APPLICATION NUMBER: U.S. 07/749,451

; FILING DATE: 23 August, 1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Heber, Sheldon O.

; REGISTRATION NUMBER: 38,179

; REFERENCE/DOCKET NUMBER: 213/006

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (213) 489-1600

; TELEFAX: (213) 955-0440

; TELEX: 67-3510

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5006 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; FEATURE:

; NAME/KEY: CDS
; LOCATION: 436...3699
; OTHER INFORMATION:
; US-08-484-565-2

alignment_scores:

Quality: 150.00 Length: 690

Ratio: 0.543 Gaps: 27

Percent Similarity: 40.000 Percent Identity: 19.565

alignment_block:

US-09-775-181-4 x US-08-484-565-2 ..

Align seg 1/1 to: US-08-484-565-2 from: 1 to: 5006

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91

829 TCAGAGCACATTCCTCTACGATTGCTGTGGGAGCACTGGCTCAGG 878

91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108

879 CGTCTCCACGCGCAGTGGCAAAATCTGCTGGGCTCTTCTACATT..... 921

108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124

922CCCCAGGTCAGTTATGGCTCCTCCAGC.....AGA 951

125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG1 141

952 CTCCTCAGCAACAGATCAATTCAGTCTTCTCCGGAACC...ATCCC 998

141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAspTrpTyrG 158

999 CAATGATGAGCAGCCAGCCACTGCCATGGCAGACATCATCGAGTATTCC 1048

158 lnAlaLeuValTrpSerLeuLeuGluGluProSerIleSer..... 172

1049 GCTGGAACTGGTGGGCAAAATTCAGCTGATGACGACTATGGGCGCGC 1098

173ArgAlaAlaIleThrPheSerT 180

1099 GGGATTGAGAAATCCGAGAGGAGCTGAGGAAGGATATCTGCATCGA 1148

180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195

1149 CTTCACTCACTCATCTCCAGTACTCTGATCAGGAAGAGATCCAGCATG 1198

196ArgGluGluSer.ArgIleLeuLeuG 204

1199 TGGTAGAGGTGATTCAAAATTCACGCCCAAGTCATCGTGTTCCTCC 1248

204 lnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuLuthr 220

1249 AGTGGCCAGATCTTGAGCCCTCATCAAGGAGATTGCCGCGCAA... 1295

221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArg 237

1296TATCAGCGGCAAGATCTGGCTGGCCAGCGGCGCTGGCCAGCT 1339

237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250

1340 CTTCCCTGATGCCATGCTCAGTACTTCCACGTCGTTGGGCGCACCAT 1389

250 rpArgArgLysaspGlyLeuGlyGlyAspLysSerHisPheLysTrpSer 266

1340 GGATTCGCTCTGAGGC.....TGGGCA 1412

267 ProProTyrLeuGlu.....Cy 272

1413 GATCCAGGCTTCGGGAATTCCTCAAGAAGGTCCATCCAGGAAGTCTG 1462

272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284

Result No.	Score	Query Match	Length	DB	ID	Description
1	136	4.3	999	2	T27628	hypothetical prote
2	133.5	4.2	1079	2	I59362	calcium/polyvalent
3	127	4.0	1078	2	A56715	calcium receptor (
4	123	3.8	1085	2	A50476	Ca(2+)-sensing rec
5	120.5	3.8	879	2	JC7160	metabotropic gluta
6	119.5	3.7	1088	2	B56715	calcium receptor (
7	116	3.6	872	2	JH0561	metabotropic gluta
8	115	3.6	1199	2	A1939	G protein-coupled
9	114	3.6	1180	2	JC2132	metabotropic gluta
10	114	3.6	1212	2	JC2131	metabotropic gluta
11	112	3.5	879	2	JH0562	metabotropic gluta
12	111	3.5	1171	2	A42916	metabotropic gluta
13	110.5	3.5	3507	2	T34513	hypothetical prote
14	109.5	3.4	868	2	JC5701	ErBB kinase activa
15	107.5	3.4	2437	2	S35611	MIB1 protein - ra
16	106.5	3.3	850	2	JC5700	ErBB kinase activa
17	106.5	3.3	860	2	JC5702	ErBB kinase activa
18	106	3.3	383	2	S53716	delta-like homeoti
19	106	3.3	551	2	T30806	metabotropic gluta
20	104.5	3.3	644	1	A0212	uromodulin precurs
21	104.5	3.3	644	2	I84634	Tamm-Horsfall prot
22	104.5	3.3	956	2	A57121	thrombospondin 3 p
23	100.5	3.1	1218	2	S71376	glutamate receptor
24	100	3.1	1993	2	T30902	sodium channel SCA
25	99.5	3.1	669	2	I38029	matrix metalloprot
26	98.5	3.1	459	2	JC5139	vitronectin precur
27	98	3.1	2531	2	T31070	notch homolog - se
28	98	3.1	2871	2	A55667	fibrillin I - bovi
29	97.5	3.0	385	2	A54785	preadipocyte facto

A: Title: Calcium sensing receptor: molecular cloning in rat and localization to nervous


```
Db 538 LTFVLVQLQ-----VFFSNCSR-----DCLAGTRKGIIEGPTCCFCEVCPDQG 581
QY 399 -----CPFCADD-----SPCFVQEDKYL-----LAIISFOGLCMMLDFVSM 435
Db 582 EYSDETDASACNKPDPFWSNHNHTSCIAKEIEFLSWTEPFGLALTLFAVLGIFLTAFLV 641
QY 436 LVVYHFRKAKSTRASGLILELTLFGSLLLYPVWILYFEPSTFRICILLRWARLLGFAIV 495
Db 642 GVPIKFRNTPIVKATNRELSYLLLFSLCCFSSSLFFIGEPQDWTCLRQPAFGISFVLC 701
QY 496 YGVTVTLKHLRVLFVLSRTAQRIPYMTGGRVVRMLAV-----ILLVFWFLGHTSSV 548
Db 702 ISCILYKTRNLVLFYFAKIPTSFHRKRWGSLNQFLVFLCTFMQIVICVIMLYTAPPSSY 761
QY 549 -CNLEKQISLI-----GQKTSDHILFNCLLI 575
762 RNOELEDIEIFTCHEGSLMALGFLIGYTCLL 793

RESULT 7
JH0561
metabotropic glutamate receptor 2 precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 13-Sep-1998
C:Accession: JH0561
R:Hanabe, Y.; Masu, M.; Ishii, T.; Shigemoto, R.; Nakanishi, S.
Neuron 8, 169-179, 1992
A:Title: A family of metabotropic glutamate receptors.
A:Reference number: JH0561; MUID:92110002
A:Accession: JH0561
A:Molecule type: mRNA
A:Residues: 1-872 <TAN>
A:Experimental source: brain
A:Comment: This protein is coupled to a G protein and evokes a variety of functions by m
C:Superfamily: metabotropic glutamate receptor 4
C:Keywords: G protein-coupled receptor; glycoprotein; phosphoprotein; transmembrane prot
F:1-18/Domain: signal sequence #status predicted <Sig>
F:19-872/Product: metabotropic glutamate receptor 2 #status predicted <MET>
F:568-590/Domain: transmembrane #status predicted <TRI>
F:605-625/Domain: transmembrane #status predicted <III>
F:637-655/Domain: transmembrane #status predicted <III>
F:680-700/Domain: transmembrane #status predicted <III>
F:726-747/Domain: transmembrane #status predicted <TRV>
F:761-782/Domain: transmembrane #status predicted <TRV>
F:795-819/Domain: transmembrane #status predicted <VII>
F:820-838/Domain: transmembrane #status predicted <VII>
F:839-856/Domain: transmembrane #status predicted <VII>
F:857-872/Domain: transmembrane #status predicted <VII>
F:203,286,338,402,547/Binding site: carboxylate (Asn) (covalent) #status predicted
F:501,675,827,837,843/Binding site: phosphate (Ser) (covalent) #status predicted
F:832/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 3.6%; Score 116; DB 2; Length 872;
Best Local Similarity 18.9%; Pred. No. 0.79;
Matches 93; Conservative 63; Mismatches 184; Indels 152; Gaps 22;
QY 141 QNKSR-----QNLQDDLDWQALWSLLE-----GEPISRAAIFTFSDLSAPAPQVF 191
Db 271 REEDARELLAATQRLNASPTWASDGGWGALESVAVGASERAAEGAITIELASPIISDFASY 330
QY 192 LOATREESRILLQDLSSAPHLANATLETWFHGLRRKWRPHLHRRGPNQGRGLGHSWR 251
Db 331 FOS-----LDPWNS-----RNPWF---REFWEERFH-----CSFR 358
QY 252 RKDLGGDKSHKFPYLECENGSKYKPGWLVTLSAIYGLQPNLVP----- 298
Db 359 ORD-----CAHSLRAVPFQESKIMFVNAVYMAHALNHMRALCPNTHLCDAMRPVN 414
QY 299 -----EPFGVMKVNDINLOKVDI-DQCSDDGWFSGTHKRLANSECMPIKGLGVLGAYE 351
Db 415 GRRLYKDFVLNFKDAPFPADTDDEVRDFREGD-----IGRN 454
QY 352 CIC-----KAGYHPGVLPVNNF-----RRGPDQHSIGSTKDYSE 387
Db 455 IFYTLRAGSGRYRYQKVGWAGELDTISFIWASPSAGPLPASRCEPCLQNEKVSQVP 514
```

```
QY 388 E---AYVCLPCREGCPF-----CADD-----SPCFVQEDKYL----- 417
Db 515 GEVCCWLCIPCO---PYEYRLDEFTCADGCLGWPNASLTGCFELPQEVIRWGDANAVGP 571
QY 418 LAIISFOGLCMMLDFVSMVWVYHFRKAKSIRASGLLILELTLFGSLLLYPVWVILYFEP 477
Db 572 VTIACLGALATL---FVLGVFVRH-NATPVVKASGRELCLYILLGGVFLCYCMTFVFIKPS 628
QY 478 TFRICILLRWARLLGFATVYGVTVTLKHLRVLYKVF---LSRTAQRIPYMT-GGRVVRMLAVI- 533
Db 629 TAVCTLRRGLGTAFSVCYSAULTTNRRIARIFGGAREGAQRPRFISPASOVAICLALIS 688
QY 534 ---LLVFWFLI 542
Db 689 GOLLIVAAWLVV 700

RESULT 8
A41939
G protein-coupled glutamate receptor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A41939; S15362
R:Hanabe, K.M.; Kujiper, J.L.; Gilbert, T.L.; Haldeman, B.A.; O'Hara, P.J.; Mulvihill
Science 252, 1318-1321, 1991
A:Title: Cloning, expression, and gene structure of a G protein-coupled glutamate rec
A:Reference number: A41939; MUID:92022526
A:Accession: A41939
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1199 <HOU>
A:Cross-references: GB:M61099; NID:g397806; PIDN:AAA19497.1; PID:g204460
A:Experimental source: cerebellum
A:Note: sequence extracted from NCBI backbone (NCBIP:60785)
R:Masu, M.; Tanabe, Y.; Tsuchida, K.; Shigemoto, R.; Nakanishi, S.
Nature 349, 760-765, 1991
A:Title: Sequence and expression of a metabotropic glutamate receptor.
A:Reference number: S15362; MUID:91156047
A:Accession: S15362
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1199 <MAS>
A:Cross-references: EMBL:X57569; NID:g56646; PIDN:CAA40799.1; PID:g56647
C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 3.6%; Score 115; DB 2; Length 1199;
Best Local Similarity 20.7%; Pred. No. 1.4;
Matches 61; Conservative 45; Mismatches 108; Indels 80; Gaps 9;
QY 357 GFYHPGVLPVNNFRRRGPDQHSIGSTKDYSE-----AYVCLPCREG- 398
Db 498 GTWHEGLVNDIDYKIQ---MKNKSGMVRVSCPECLKGQIKVKGEVSCWCTTACKNE 554
QY 399 -----CPFC-----ADDSPCFVQEDKYL-----LAIISFOGLCMMLDFVSMV 438
Db 555 FYQDEFTCRACDLGWPNNAELTGCEPIPVRYLEWSDIESIIAIAFSCIGILVTLFVTLIF 614
QY 439 YHFRKAKSTRASGLLILELTLFGSLLLYPVWVILYFEPSTFRICILLRWARLLGFAIVYT 498
Db 615 VLYRDTVPVVKSSRELCLYIILAGIFLVYVCPFTLIAKPTTTCYLQRLVLGLSSAMCYSA 674
QY 499 VTLKHLRVLV-----FLSRTAQRIPYMTGGRVVRMLAVITLLVFWFLG 544
Db 675 LVYTKNTRTARILAGSKKICTRKPRFMSAWAOVITIASILISVQLTLVLTLLI----- 726
QY 545 TSSVCNLEKQISLIGCKQKTSQD-HLIFNNCLIDRWDMYTAAGWMSLVSYDGLTI 597
Db 727 -----MEPPMPILSYPSIKEVYLICN-----TSNLGWVAPVGYNGLLI 764

RESULT 9
```

```
F:644-664/Domain: transmembrane #status predicted <TM3>
F:694-714/Domain: transmembrane #status predicted <TM4>
F:738-759/Domain: transmembrane #status predicted <TM5>
F:773-794/Domain: transmembrane #status predicted <TM6>
F:803-827/Domain: transmembrane #status predicted <TM7>
```

Query Match	3.6%	Score 114;	DB 2;	Length 1212;
Best Local Similarity	20.4%	Pred No 1	7.	

[illegible]

```

372  RGPDQHI---SGSTKDVSEE-----AYVCLPCREG-----CPFC 402
      :       :       :       :
493  KMDDDVSWKSKNIIIRVSCSEPCFKGOTKVRKGEVSCCCTCTPCKENYEVFDEVYTCRAC 552
      :       :       :       :

```

403 QY -----ADD-SPCFQVEDKYLR-----LATISFQGLCMLLDVFSMLVVHYHFRKAKSIRA 449
||| : ||| :
553 OLGSWPTDDELTCGDLIPVOYLRWGDPEPIAAVVFACLGLLATLEVTVWFIIYRTPPVKKS 612
||| : ||| : ||| : ||| : ||| :

450 SGLILETILFSGSLLLYFPVVILYFEPSTFRCLLRNARLGFATVYGTVTLKLRHLVKV 509 QY
613 SSRELICYIIILAGICLGYLCTFCLIAKPKOICYVLORTIGIGLSPANSYSAVTKYKNRIAR 572 Db

Qy	510	FLSRTAQRI-----PYMTGGVRMRMLAVILLV	537
		: :	
		: :	
Db	673	-LAGSKKIKCTKKPREMSACQLVIAFILI	703

RESULT 11
JH0562

C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 13-Sep-1998
C:Accession: JH0562

Neuron 8, 169-179, 1992
A.Title: A family of metabotropic glutamate receptors.
A.Reference number: JH0561; MUID:92110002

A: Molecule type: mRNA
A: Residues: 1-879 <TAN>
A: Experimental source: brain

C:Superfamily: metabotropic glutamate receptor 4
C:Keywords: G protein-coupled receptor; glycoprotein; phosphoprotein; transmembrane protein
F1-22/Ydomain: signal sequence #status predicted <SIG> 22-2970/

F.577-599/Domain:	transmembrane	#status	predicted
F.614-634/Domain:	transmembrane	#status	predicted
F.646-664/Domain:	transmembrane	#status	predicted
F.690-700/Domain:	transmembrane	#status	predicted

F:735-756/Domain:	transmembrane	#status predicted
F:770-791/Domain: <th>transmembrane</th> <th>#status predicted</th>	transmembrane	#status predicted
F:804-828/Domain: <th>transmembrane</th> <th>#status predicted</th>	transmembrane	#status predicted
F:209-362/414-430/Binding site: <th>carbohydrate (Asn)</th> <th>#status predicted</th>	carbohydrate (Asn)	#status predicted

Query Match 3 5% Score 112.00 Length 879.
 F:010,845/Binding site: pnosnate (Ser) (covalent) #status predicted

Best Local Similarity 22.6%; Pred. NO. 1.0;
Matches 42; Conservative 30; Mismatches 76; Indels 38; Gaps 7;

Qy	390	YVCLPCRE-----GCPCF-----ADSPCFVOEDKYLR-----LATISF 423
		: : : : : : : : : : : : : :
Db	529	WICIPCEPYEVLVDEFTCMDCGQCWPTDLSGCYNLPEDYIKWEDAWAIGPVTIACLG 588
		: : : : : : : : : : : : : :
Qy	424	OGLCMLLDFVSMVLVYHFRKAKSIRASGLILLETILFGSLILYFPVILYFSPFCRIL 483
		: : : : : : : : : : : : : : : : : : : : : :
Db	589	LCTCIVL-----TVTFKHNNTPLVKASGRELCYILFGVSLCYCMFTFFIAKPSVICAL 643
		: : : : : : : : : : : : : : : : : : : : : :
Qy	484	LRMARLLGFATVYGTVTLKLRHVLKVF--LSRTAQRIPYMT--GGRVMRMIAVIL---LV 536
		: : : : : : : : : : : : : : : : : : : : : :
Db	644	RRUGLGTSPAICYALLTKNTCIARFDGVKNQAQRPKFISFSSQVTCGLGUILVQIVMW 703
		: : : : : : : : : : : : : : : : : : : : : :
Qy	537	VFWFLI 542
		: : : : : : : : : : : : : : : : : : : : : :
Db	704	SVWLIL 709
		: : : : : : : : : : : : : : : : : : : : : :

SULT 12
A42916
metabotropic glutamate receptor mGluR5 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A42916
R:Abe, T.; Sugihara, H.; Nawa, H.; Shigemoto, R.; Mizuno, N.; Nakanishi, S.
J. Biol. Chem. 267, 13361-13368, 1992
A:Title: Molecular characterization of a novel metabotropic glutamate receptor mGluR5 co
A:Reference number: A42916; MUID:92317054
A:Accession: A42916
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1171 >ABE>
A:Cross-references: GB:D10891; NID:g220813; PIDN:BAA01711.1; PID:d1002186; PID:g220814
A:Experimental source: brain
A>Note: sequence extracted from NCBI backbone (NCBI:107749, NCBI:P:107750)
C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 3.5%; Score 111; DB 2; Length 1171;
Best Local Similarity 20.8%; Pred. No. 2.8;
Matches 80; Conservative 54; Mismatches 171; Indels 80; Gaps 17;

QY	219	ETEWFCILRRKWRP	PHLLRRCPN	QPGRLGHSNR	-----RKDGLGDKSHFKWSPPYLECE	273
Db	332	DVKWFDDYY	LKLRPE	TNLNRP	-----WFOEFWOHRFQCLEGFAENSKYNT	381
Db	274	NG	-----SYKPCWLV	-TJSSALYGL-	-----QPNLVPBERGV	311
Db	382	SSLTRTHHVQDS	KMGFVNAI	YSMAYGLHNQMSL	CPYAGLCDAKPIDGRKLLDSLM	441
QY	312	KVDIDOCSSD	GWFSGTHK	CHLNNSCEPIK	GLGVILGAYECTKAGFYHPGVLPV	366
Db	442	KTNFTGVSGDM	ILFDENGSD	SPGRYEIMNFKENG	---KDYFDYINVSGNDGELKMDDDDEV	498
QY	367	-----NFRRRGPD	QHI	-SGSTKDVSEE	-----NYVLCPREG	403
Db	499	WSKKNKNI	IRSVCEPCEK	GQIKVIRKEVSC	WCCTTPCKENEYVFDEYTCRACOLGSWPT	558
QY	404	DD-SPCFVQED	KYLR	-----LAIISFOGLCMLL	DFVSMVLVYHRRKAKSI	456
Db	559	DDUTGOLIPVQ	YLRWGDEP	PTAAVFAFCGLL	LATLVTVITFIYDRTPVKXSSRELCY	618
QY	457	TIILFGLSLLY	FPVVIYF	FPSTFRCLLRW	ALLGFATVYGTUULKLRHLVKVFLSR	516
Db	619	IILAGICLGY	LCTFCLIAK	PQIYCYLQ	RIGIGLSPAMSYSAVTKTNRIARI	677
QY	517	RI	-----PYWTGGVR	MRMLAVILLV	537	
Db	678	KICTKPR	PREMSACAQ	LVI	AFILICI	702

RESULT 13
T34513

hypothetical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Favello, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid zK783.
A:Reference number: Z21536
A:Accession: T34513
A:Status: preliminary; translated from CB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:zK783.1
A:Experimental source: strain Bristol N2; clone zK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/1
3504/1

Query Match 3.5%; Score 110.5; DB 2; Length 3507;
Best Local Similarity 18.0%; Pred. NO. 12;
Matches 76; Conservative 51; Mismatches 137; Indels 159; Gaps 19;

QY 20 GAVGASRDPQGRPDSP-----RERTPKG-----KPHAQQPGRA 52

Db 1227 GDNGETSGVDGKPTTPAPTPSSSAESSTSRIPPTTSEASPEGGSGEAGVPESP DGSGESST 1286

QV 53 SASDSSAPWSRSTDGTILAOKLAEVPMVDVASLYLTGDSHOLKRANCSGRYELAGLPQKW 112

db 1287 SAPDGVSP TSSAT-----APEVPTTSAS--STPDA-----VFESGIPS-- 1322

0v 113 PALASAHPSIHRALDTI,THATNEI NVMIOSNKSBEONI,ODDIDWYQAI,VWSI,LECEPSIS 172

```

Dh 1323 TSKBPTAPP-----LETTAPSTFVTSDF--ECSCTEFESTI DDT---ECSCTEFESTTSSADPME 1371

```

173 PRAATTESTOISADAPDOVEIANTDEESPTIYONI SSSADUI ANAETI EMELEUCI DDVWDD 333

DB	1373	PATVI	DBONP	NEW	DEP	-----	TKD	DAY	DBONP	TC	ADON	NO	OSVE	-----	1410
		:	:	:			:	:	:		:	:	:		

[illegible][illegible][illegible]

DATE	DESCRIPTION	AMOUNT	CHECK	BALANCE
10/1/54	10/1/54	100.00		100.00
10/2/54	10/2/54	100.00		200.00
10/3/54	10/3/54	100.00		300.00
10/4/54	10/4/54	100.00		400.00
10/5/54	10/5/54	100.00		500.00
10/6/54	10/6/54	100.00		600.00
10/7/54	10/7/54	100.00		700.00
10/8/54	10/8/54	100.00		800.00
10/9/54	10/9/54	100.00		900.00
10/10/54	10/10/54	100.00		1000.00
10/11/54	10/11/54	100.00		1100.00
10/12/54	10/12/54	100.00		1200.00
10/13/54	10/13/54	100.00		1300.00
10/14/54	10/14/54	100.00		1400.00
10/15/54	10/15/54	100.00		1500.00
10/16/54	10/16/54	100.00		1600.00
10/17/54	10/17/54	100.00		1700.00
10/18/54	10/18/54	100.00		1800.00
10/19/54	10/19/54	100.00		1900.00
10/20/54	10/20/54	100.00		2000.00
10/21/54	10/21/54	100.00		2100.00
10/22/54	10/22/54	100.00		2200.00
10/23/54	10/23/54	100.00		2300.00
10/24/54	10/24/54	100.00		2400.00
10/25/54	10/25/54	100.00		2500.00
10/26/54	10/26/54	100.00		2600.00
10/27/54	10/27/54	100.00		2700.00
10/28/54	10/28/54	100.00		2800.00
10/29/54	10/29/54	100.00		2900.00
10/30/54	10/30/54	100.00		3000.00
10/31/54	10/31/54	100.00		3100.00
11/1/54	11/1/54	100.00		3200.00
11/2/54	11/2/54	100.00		3300.00
11/3/54	11/3/54	100.00		3400.00
11/4/54	11/4/54	100.00		3500.00
11/5/54	11/5/54	100.00		3600.00
11/6/54	11/6/54	100.00		3700.00
11/7/54	11/7/54	100.00		3800.00
11/8/54	11/8/54	100.00		3900.00
11/9/54	11/9/54	100.00		4000.00
11/10/54	11/10/54	100.00		4100.00
11/11/54	11/11/54	100.00		4200.00
11/12/54	11/12/54	100.00		4300.00
11/13/54	11/13/54	100.00		4400.00
11/14/54	11/14/54	100.00		4500.00
11/15/54	11/15/54	100.00		4600.00
11/16/54	11/16/54	100.00		4700.00
11/17/54	11/17/54	100.00		4800.00
11/18/54	11/18/54	100.00		4900.00
11/19/54	11/19/54	100.00		5000.00
11/20/54	11/20/54	100.00		5100.00
11/21/54	11/21/54	100.00		5200.00
11/22/54	11/22/54	100.00		5300.00
11/23/54	11/23/54	100.00		5400.00
11/24/54	11/24/54	100.00		5500.00
11/25/54	11/25/54	100.00		5600.00
11/26/54	11/26/54	100.00		5700.00
11/27/54	11/27/54	100.00		5800.00
11/28/54	11/28/54	100.00		5900.00
11/29/54	11/29/54	100.00		6000.00
11/30/54	11/30/54	100.00		6100.00
12/1/54	12/1/54	100.00		6200.00
12/2/54	12/2/54	100.00		6300.

[illegible]

DD 1477 ECF10F KRADD GSCQDIDE CIEHNSIC CGANACVNRK 1314

QY 413 DKY 413

DB 1513 GY 131/

RESULT 14

JC5701 : ErbB kinase activator alpha1. brain and thymus - rat

C;Species: Rattus norvegicus (Norway rat)
C;Date: 25-Nov-1997 #sequence revision 25-Nov-1997 #text change 11-Jan-2000

C;Accession: JC5701; PC4411
R.Higashiyama S : Horiyawa M : Yamada K : Ichino N : Nakano N : Nakayama

J. Biochem. 122, 675-680, 1997

A;Reference number: JC5700; MUID:98006324

A;Molecule type: mRNA

A;Cross-references: DDBJ:D89995; NID:g2605629; PIDN:BAA2344.1; PID:g260563

A:Molecule type: protein
A:Residues: 128-162 <HI2>
A:Experimental source: PC-12 cell
C:comment: This protein is a member of the epidermal growth factor family. It is functional in the differentiation of MDA-MB-453 cells.
C:Superfamily: unassigned EGF-related proteins; EGF homology
F:361-397/Domain: EGF homology <EGF>

Query Match 3.4%; Score 109.5; DB 2; Length 868;
Best Local Similarity 19.1%; Pred. No. 2.5;
Matches 92; Conservative 57; Mismatches 173; Indels 159; Gaps 20;

Qy 31 PDSPRETPKPKHAQOP-----GRASAS-----DSAPHSRSTGDTILA----- 71
Db 74 RPAAPPEPRPQPPRPAARAAARSAAAGMRDPAGGSMILFGVSLACYSPSL 133
Qy 72 ---QKLAPEPMDVASYLY-----TGDSHQLKRANCSGRVELAGLCKWPALASAHPSL 122
Db 134 KSVQDOAYKAPVVVEGKVOGLAPAGGSSNSTREPPASGRVALYKVLDKWP----- 184
Qy 123 HRALDTLTHATNFLNVLQSNKSRQNLQ----DDLWYQALVMSLLEGEPSISRAAITF 178
Db 185 -----LRSGGLREQVTSVSGCAPLERNQRYIFEL---EP--TEQPLVF 223
Qy 179 STDLSAPAPQVFLQATREESRIILQDLSSSAPHLANATLETWFHGLRRKWRPHLHRRG 238
Db 224 KT--AFAPVDPNGKNIKKEVGKILCTDCAT-----RPKLKKMK 259
Qy 239 PNQPRGLGHSMRRKDGGLGGDKSHFKPPVYLECENGSKYKQWLVTLSAI---YG----- 291
Db 260 SQTVEGKEQSLKCAAGAGNPQPSRW-----FKDGKEINRSDIRIKYGNRK 308
Qy 292 ---LQPNLVP-----EPRGMKVDINLQKVIDQCSS--DGWFSQTHKCH----- 331
Db 309 NSRLQFNKVKVEDAGEYVCEANILGKDTVGRCHVNSVSTLSSWSCHARKCNETAKSY 368
Qy 332 -LNNSCEPIKGLFVLCAYCEICKAGYHPCVLPVNNRRRGPD---QHISGSTKQVSE 387
Db 369 CVNGGVCIYIEG-----INQLSKCPNGFGQRCLEKPLRLYMPDPKQKHLGFELKE-AE 423
Qy 388 EAYVCLPCREGPCFCAADSPCFQVEDKYLRLAITSFOGLCMLODFVSMVYVHPRKAKSI 447
Db 424 ELY-----QKRVLITICVALLVGVGVVAYCKTRKQ 457
Qy 448 R 448
Db 458 R 458

RESULT 15

3611
A: protein - rat
A:Alternate names: angiotensinogen gene-inducible enhancer-binding protein; c-myc intron
C:Species: Rattus norvegicus (Norway rat)
C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 20-Jun-2000
C:Accession: S53611; A39796; S22292; I58280
R:Makino, R.; Akiyama, K.; Yasuda, J.; Mashiyama, S.; Honda, S.; Sekiya, T.; Hayashi, K.
Nucleic Acids Res. 22, 5679-5685, 1994
A:Title: Cloning and characterization of a c-myc intron binding protein (MIBP1).
A:Reference number: S53611; MUID:95140632
A:Accession: S53611
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-2437 <NAK>
A:Cross-references: EMBL:D37951; NID:g1408559; PIDN:BAA07168.1; PID:g662296
R:Ron, D.; Brasier, A.R.; Habener, J.F.
Mol. Cell. Biol. 11, 2887-2895, 1991
A:Title: Angiotensinogen gene-inducible enhancer-binding protein 1, a member of a new family of transcription factors.
A:Reference number: A39796; MUID:91203912
A:Accession: A39796
A:Status: preliminary
A:Molecule type: mRNA

A:Residues: 'I', 1523-2263, 'K', 2265-2437 <RON>
A:Cross-references: GB:M65251; NID:g202790; PIDN:AAA40698.1; PID:g202791
R:Mitchellmore, C.; Traboni, C.; Cortese, R.
Nucleic Acids Res. 19, 141-147, 1991
A:Title: Isolation of two cDNAs encoding zinc finger proteins which bind to the alpha
A:Reference number: I58280; MUID:91187610
A:Accession: S22292
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 'E', 1729, 'R', 1734-2437 <MIT>
A:Cross-references: EMBL:X54249; NID:g57517; PIDN:CAA38150.1; PID:g57518
A:Note: the authors did not translate the codon for residue 1
C:Superfamily: HIV-E2 enhancer-binding protein
C:Keywords: DNA binding; duplication; metal binding; transcription regulation; zinc f
F:191-211/Region: zinc finger CCHH motif
F:219-241/Region: zinc finger CCHH motif
F:934-940/Region: nuclear location signal
F:947-979/Region: serine-rich
F:1792-1812/Region: zinc finger CCHH motif
F:1820-1842/Region: zinc finger CCHH motif
F:1890-1917/Region: acidic

Query Match 3.4%; Score 107.5; DB 2; Length 2437;
Best Local Similarity 21.1%; Pred. No. 13;
Matches 71; Conservative 54; Mismatches 133; Indels 79; Gaps 15;

Qy 18 GLGAVGSRDPQGRDPSRRTPKGPHAQOPGRASADSSAPWSKSTGDTILAOKLAE 77
Db 1278 GASGLHKNLPKPFSDPGSKSTEAPTEQLLRDFASENACP-LQSLPGTVVPRIQTH 1336
Qy 78 VPMDVASYLYTGDSHQLKRANCSGRYELAGLGPALASAHPSLHRAIDTLTHATNFIN 137
Db 1337 VP-SYGSVMVTSISQILGONS-----PAIV-----ICKVDENMTQRTLVN 1376
Qy 138 VMLQS---NKSRQNLQDLDWYQALVMSLLEGE-----SISRAAITFST 180
Db 1377 AAMOGIGFNIAOVLGQRTGLEKYP--LWKVPQTLPLGLESIPLCLPSTDSAAISGSK 1434
Qy 181 DLSAPAPQVFLQATREESRI-----LQDLSSSAPHLANATLETWFHGLRRKWRP 232
Db 1435 RMLSPASSLELFEMETKQKRKVEEKMYQIVVEL--SAVELTNSDIK-----KLSRPQKP 1488
Qy 233 HLHRRGNQGPRLGHSWRRKDG-----GGDKSHFKWS---PPYLECENGSKYKPGWLVT 284
Db 1489 QLVROGCASEPKD-GSSQSRSFSSLSFSSSQDHPAASGPPPPNREILSGSRAP----- 1542
Qy 295 LSSAIYGLQPNLVEFRGVMKVDINLQKVIDQCSSD 321
Db 1543 -----PR--RKFSGPSRESSELDIDETSSD 1568

Search completed: November 1, 2001, 10:15:15
Job time: 813 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 10:20:45 ; Search time 24.77 Seconds
(without alignments)
828.383 Million cell updates/sec

Title: US-09-775-181-4

Perfect score: 3199

Sequence: 1 MCAMAYPLLLCLLLAQLGLG.....YMTAVGMWSLVSVDGLTIFQ 599

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	161.5	5.0	976	1 MGR_DROME	P91685 drosophila
2	136	4.3	999	1 MGR1_CABEL	Q09630 caenorhabdi
3	134.5	4.2	1079	1 CASR_MOUSE	Q99y96 mus musculu
4	133.5	4.2	1079	1 CASR_RAT	P48442 rattus norv
5	127	4.0	1078	1 CASR_HUMAN	P41180 homo sapien
6	123	3.8	1085	1 CASR_BOVIN	P35384 bos taurus
7	120	3.8	872	1 MGR2_HUMAN	Q14416 homo sapien
8	116	3.6	872	1 MGR2_RAT	P31421 rattus norv
9	116	3.6	1194	1 MGR1_HUMAN	Q13255 homo sapien
10	115.5	3.6	877	1 MGR3_HUMAN	Q14832 homo sapien
11	115	3.6	1199	1 MGR1_RAT	P23385 rattus norv
12	114	3.6	1212	1 MGR5_HUMAN	P41594 homo sapien
13	112	3.5	879	1 MGR3_RAT	P31422 rattus norv
14	111	3.5	1203	1 MGR5_RAT	P31424 rattus norv
15	109.5	3.4	868	1 MGR2_RAT	Q35569 rattus norv
16	106.5	3.3	850	1 MGR2_HUMAN	Q14511 homo sapien
17	106	3.3	383	1 DLK_HUMAN	P80370 homo sapien
18	104.5	3.3	644	1 UROM_RAT	P27590 rattus norv
19	104.5	3.3	956	1 TSP3_HUMAN	P49746 homo sapien
20	101	3.2	912	1 MGR4_HUMAN	Q14833 homo sapien
21	100	3.1	356	1 DCUP_DROME	Q9V595 drosophila
22	99	3.1	669	1 MM15_HUMAN	P15111 homo sapien
23	98.5	3.1	459	1 VTNC_PIG	P48819 sus scrofa
24	98	3.1	2871	1 FBNI_BOVIN	P98133 bos taurus
25	97.5	3.0	749	1 COA2_PAVEN	P22964 porcine par
26	97.5	3.0	2871	1 FBNI_HUMAN	P35555 homo sapien
27	97	3.0	915	1 MGR7_HUMAN	Q14831 homo sapien
28	97	3.0	1078	1 S24A_HUMAN	Q95486 homo sapien
29	97	3.0	2911	1 FBNI_HUMAN	P35556 homo sapien
30	96.5	3.0	986	1 YHDP_ECOLI	P46474 escherichia
31	96.5	3.0	2871	1 FBNI_MOUSE	Q61554 mus musculu
32	95	3.0	2907	1 FBNI_MOUSE	Q61555 mus musculu
33	94.5	3.0	956	1 TSP3_MOUSE	Q05895 mus musculu

ALIGNMENTS

RESULT 1

ID	MGR_DROME	STANDARD;	PRT;	976 AA.
DT	P91685;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	METABOTROPIC GLUTAMATE RECEPTOR PRECURSOR.			
GN	GLYRA OR GLU-RA.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1];			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=OREGON-R;			
RX	MEDLINE=96421661; PubMed=8824309;			
RA	Parmentier M.L., Pin J.P., Bockaert J., Grau Y.;			
RT	"Cloning and functional expression of a Drosophila metabotropic			
RT	glutamate receptor expressed in the embryonic CNS.;"			
RL	J. Neurosci. 16:6687-6694(1996).			
CC	-1- FUNCTION: RECEPTOR FOR GLUTAMATE.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.			
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED IN THE CNS OF THE LATE EMBRYO.			
CC	-1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL: X99675; CAA67993.1; -			
DR	GDDB; GCR_1123;			
DR	FlyBase; FBgn0019985; Glu-RA.			
DR	InterPro; IPR000162; -			
DR	InterPro; IPR000337; -			
DR	InterPro; IPR001828; -			
DR	Pfam; PF000003; 7tm_3; 1.			
DR	Pfam; PF01094; ANF_receptor; 1.			
DR	PRINTS; PR00248; GPCRMR.			
DR	PRINTS; PR00593; MTABOTROPICR.			
DR	PROSITE; PS00979; G_PROTEIN_RECF_F3.1; 1.			
DR	PROSITE; PS00980; G_PROTEIN_RECF_F3.2; 1.			
DR	PROSITE; PS00981; G_PROTEIN_RECF_F3.3; 1.			
DR	PROSITE; PS02529; G_PROTEIN_RECF_F3.4; 1.			
KW	G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.			
FT	SIGNAL 1 ?			
FT	CHAIN ? 976			
FT	DOMAIN ? 626			
FT	DOMAIN 627 649			
FT	TRANSMEM 650 663			
FT	DOMAIN 654 684			
FT	TRANSMEM 664			

34	94	2.9	381	1	OPS2_DROME	P08099 drosophila
35	94	2.9	915	1	MGR7_RAT	P35400 rattus norv
36	94	2.9	1403	1	NID2_MOUSE	O88322 mus musculu
37	93.5	2.9	385	1	DLK_MOUSE	Q09163 mus musculu
38	93.5	2.9	621	1	KNHL_BOVIN	P01044 bos taurus
39	93.5	2.9	808	1	FTFB_DROME	O05192 drosophila
40	93.5	2.9	840	1	SYL_BORBU	O01267 borrelia bu
41	93.5	2.9	1133	1	EGF_RAT	P07522 rattus norv
42	93	2.9	810	1	NEL1_HUMAN	Q92832 homo sapien
43	93	2.9	908	1	MGR8_MOUSE	P47743 mus musculu
44	92.5	2.9	810	1	NEL1_RAT	O62919 rattus norv
45	92.5	2.9	1184	1	FBL2_HUMAN	P98095 homo sapien

01-OCT-2000 (Rel. 40, Last annotation update)
EXTRACELLULAR CALCIUM-SENSING RECEPTOR PRECURSOR (CASR) (PARATHYROID
CELL CALCIUM-SENSING RECEPTOR).
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
STRAIN=C57BL/6; TISSUE=Kidney;
MEDLINE=20092890; PubMed=10625662;
Oda Y., Tu C.-L., Chang W., Crumrine D., Koemueves L., Mauro T.,
Elias P.M., Bikie D.D.;
"The calcium sensing receptor and its alternatively spliced form in
murine epidermal differentiation.";
J. Biol. Chem. 275:1183-1190(2000).
[2]
SEQUENCE FROM N.A. (ISOFORM A), AND TISSUE SPECIFICITY.
STRAIN=BLAK SWISS X 129/SVJ; TISSUE=Kidney;
MEDLINE=20119279; PubMed=10652312;
Pi M., Garner S.C., Flannery P., Spurney R.F., Quarles L.D.;
"Sensing of extracellular cations in CasR-deficient osteoblasts.
Evidence for a novel cation-sensing mechanism.";
J. Biol. Chem. 275:3256-3263(2000).
[3]
SEQUENCE OF 256-600 FROM N.A. (ISOFORM A), AND TISSUE SPECIFICITY.
TISSUE=Epiphyseal cartilage;
MEDLINE=20043955; PubMed=10579354;
Chang W., Tu C., Chen T.-H., Koemueves L., Oda Y., Pratt S.A.,
Miller S., Shoback D.;
"Expression and signal transduction of calcium-sensing receptors in
cartilage and bone.";
Endocrinology 140:5883-5893(1999).
[4]
SEQUENCE OF 507-582 FROM N.A. (ISOFORM A).
STRAIN=NMRI; TISSUE=Brain;
Hildenbrand J., Ammon H.P.T., Wahl M.A.;
Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
[5]
SEQUENCE OF 562-814 FROM N.A.
TISSUE=Kidney;
Moawad T.I., Riccardi D.;
Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[6]
SEQUENCE OF 646-799 FROM N.A., AND TISSUE SPECIFICITY.
MEDLINE=97231187; PubMed=9076582;
Quarles L.D., Hartle J.E. II, Siddhanti S.R., Guo R., Hinson T.K.;
"A distinct cation-sensing mechanism in MC3T3-E1 osteoblasts
functionally related to the calcium receptor.";
J. Bone Miner. Res. 12:393-402(1997).
-!- FUNCTION: SENSES CHANGES IN THE EXTRACELLULAR CONCENTRATION OF
CALCIUM IONS. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A
G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
MESSENGER SYSTEM.
-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
-!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE
PRODUCED BY ALTERNATIVE SPLICING.
-!- TISSUE SPECIFICITY: EPIDERMIS, KIDNEY AND CARTILAGE.
-!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL; AF110178; AAD28371.1; -;
EMBL; AF110179; AAD28372.1; -;
EMBL; AF128842; AAD40638.1; -;
EMBL; AF068900; AAC19388.1; -;

DR EMBL; AB027140; BAA77688.1; -;
DR EMBL; AF002015; AAC53252.1; -;
DR EMBL; AF159565; AAF00193.1; -;
DR MGD; MGI:1351351; Gprc2a.
DR InterPro; IPR000068; -;
DR InterPro; IPR000337; -;
DR InterPro; IPR001828; -;
DR Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMR.
DR PRINTS; PR00592; CASENINGR.
DR PROSITE; PS00979; G_PROTEIN_RECP_F3_1; 1.
DR PROSITE; PS00980; G_PROTEIN_RECP_F3_2; 1.
DR PROSITE; PS00981; G_PROTEIN_RECP_F3_3; 1.
DR PROSITE; PS00982; G_PROTEIN_RECP_F3_4; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Alternative splicing.
FT SIGNAL 1 19 POTENTIAL
FT CHAIN 20 1079 EXTRACELLULAR CALCIUM-SENSING RECEPTOR.
FT DOMAIN 20 612 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 613 635 I (POTENTIAL).
FT DOMAIN 636 649 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 650 670 II (POTENTIAL).
FT DOMAIN 671 681 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 682 700 III (POTENTIAL).
FT DOMAIN 701 724 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 725 745 IV (POTENTIAL).
FT DOMAIN 746 769 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 770 792 V (POTENTIAL).
FT DOMAIN 793 805 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 806 828 VI (POTENTIAL).
FT DOMAIN 829 836 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 837 862 VII (POTENTIAL).
FT DOMAIN 863 1079 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 386 386 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 446 446 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 468 468 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 488 488 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 541 541 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 461 537 MISSING (IN ISOFORM B).
FT CONFLICT 45 45 A -> S (IN REF. 2).
FT CONFLICT 304 304 L -> P (IN REF. 3).
FT CONFLICT 410 410 G -> D (IN REF. 2 AND 3).
FT CONFLICT 566 566 V -> A (IN REF. 2).
FT CONFLICT 595 595 Y -> H (IN REF. 2, 3 AND 5).
FT CONFLICT 610 610 E -> V (IN REF. 5).
FT CONFLICT 814 814 F -> L (IN REF. 5).
FT CONFLICT 889 889 L -> I (IN REF. 2).
FT CONFLICT 906 909 TGSN -> SGWI (IN REF. 2).
FT CONFLICT 1057 1057 V -> M (IN REF. 2).
FT CONFLICT 1064 1064 V -> A (IN REF. 2).
FT CONFLICT 1076 1076 I -> V (IN REF. 2).
SQ SEQUENCE 1079 AA; 120839 MW; AAF8D8D472736D6E CRC64;

Query Match 4.2%; Score 134.5; DB 1; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.012;
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;

QY 235 HRRGNPGRLGHGSHWRKDLGGDKSHFKWSPYBCEGNSYKPGWLVTLSAIYGLQP 294
Db 377 HEEG---GNRLNLSSTAFLPRLCTGDENINSVETPYMEHLIRISYNYLVAYISTAHALQD 433
QY 295 --NLVPEFRGVKMDINLQKVDIDQCSGSDGWFSGTHKCHLNNS-----ECMPIKG 342
Db 434 IYTCPLG-RGLF---TNGSCADIKV--EAQVVKLHLRLNFTNNMGEOVTFDECGDLVG 487
QY 343 LGFVL-----GAYECICKAG---FYHPGVLPVNNFRRRGPDQHS----- 379

```

Db 488 NYSIINWHLSPEDGSIVFKEGVYINVYAKKGERLFINEGKILWSGFSREVFFSCSRDQ 547
QY 380 -GSTKDVSEB-----AYVCLPREG-----CPFCADD-----SPCFVQEDKYL 417
Db 548 AGTRKGIIEGPTCCFECVCEPDGEYSGETDASACDKCPDFHSNENYTSIAKEIEFLA 607
QY 418 -----LAISFOGLCMLDFVSMVYVYHFRKAKSIRASGLILLETILFGSLLLYFPVVI 471
Db 608 WTEPFGIALTFLAVLGIFLTAFLGVFIKFRNTPIVKATNRELSYLLLSLCCFSSSLF 667
QY 472 LYEPSTFCILLRWALLGCFATVYGTVLKLRVLKVPFLSRTAQRIPIYMTGGRVWRMLA 531
Db 668 FIEGPDQWTCRLRQPAFGISFVLCISILVKTNRVLVFEAKIPTSFHRKMWGLNQFL 727
QY 532 V-----ILLVFWFLIGTSSVCNLEKQISLIGOGKTSOHLINMCLIDRWDMYMTAV 584
Db 728 VFICTFMQIICIIWLYTAPPSSY-RNHELE-----DELIITC-----HEGSLMAL 773
QY 585 GMSLVSY 592
Db 774 G--SLIGY 779

RESULT 4
CASR_RAT ID CASR_RAT STANDARD; PRT; 1079 AA.
AC P48442;
DC 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE EXTRACELLULAR CALCIUM-SENSING RECEPTOR PRECURSOR (CASR) (PARATHYROID
DE CELL CALCIUM-SENSING RECEPTOR).
GN CASR OR PCAR1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Kidney outer medulla;
RX MEDLINE=95116508; PubMed=7816802;
RA Riccardi D., Park J., Lee W., Gamba G., Brown E.M., Hebert S.C.;
RT "Cloning and functional expression of a rat kidney extracellular
RT calcium/polyvalent cation-sensing receptor.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:131-135(1995).
RN [2]
RP SEQUENCE OF 1-294 FROM N.A.
RC STRAIN=WISTAR;
RX MEDLINE=95241465; PubMed=7724534;
RA Ruat M., Showman A.M., Snyder S.H.;
RT "Calcium sensing receptor: molecular cloning in rat and localization
RT to nerve terminals.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3161-3165(1995).
CC -!- FUNCTION: SENSE CHANGES IN THE EXTRACELLULAR CONCENTRATION OF
CC CALCIUM IONS. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A
CC G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
CC MESSENGER SYSTEM.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U10354; AAC52149.1; -.
CC EMBL; U20289; AAC52195.1; -.
CC GCRDb; GCR_1449; -.
CC InterPro; IPR000068; -.

```

```

DR InterPro; IPR000337; -.
DR InterPro; IPR001828; -.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRGR.
DR PRINTS; PR00592; CASSENSINGR.
DR PROSITE; PS00379; G-PROTEIN_RECEP_F3_1; 1.
DR PROSITE; PS00980; G-PROTEIN_RECEP_F3_2; 1.
DR PROSITE; PS00981; G-PROTEIN_RECEP_F3_3; 1.
DR PROSITE; PS0259; G-PROTEIN_RECEP_F3_4; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 1079
FT DOMAIN 20 612
FT TRANSMEM 613 635
FT DOMAIN 636 649
FT TRANSMEM 650 670
FT DOMAIN 671 681
FT TRANSMEM 682 700
FT DOMAIN 701 724
FT TRANSMEM 725 745
FT DOMAIN 746 769
FT TRANSMEM 770 792
FT DOMAIN 793 805
FT TRANSMEM 806 828
FT DOMAIN 829 836
FT TRANSMEM 837 862
FT DOMAIN 863 1079
FT CARBOHYD 90 90
FT CARBOHYD 130 130
FT CARBOHYD 261 261
FT CARBOHYD 287 287
FT CARBOHYD 386 386
FT CARBOHYD 446 446
FT CARBOHYD 468 468
FT CARBOHYD 488 488
FT CARBOHYD 541 541
FT CARBOHYD 594 594
FT SEQUENCE 1079 AA; 120867 MW; D7664550361F9736 CRC64;

Query Match 4.2%; Score 133.5; DB 1; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.014;
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;

QY 235 HRRGPNQGRGLGHWRKDKGLGDKSHKFPKPPVLECEGNSYKPGWLVTLSAIVGLQ 294
Db 377 HEEG---GNRLNLSSTAFRPLCTGDNINSVETPTMDYEHLSYNYLVAVISIAHALQD 433
QY 295 --NLVPEFRGMKVDINLOKVDIQDCSSDGWFSGTHKCHLNN-----ECMPIKG 342
Db 434 IYTCPLP-RGLF--TNGSCADIKV--BAQVVKLRLHRLFTNNMGEQVTFDECDLVG 487
QY 343 LGFVL-----GAYECICKAG---FYHPGVLPVNNFRRRGPDQHS----- 379
Db 488 NYSIINWHLSPEDGSIVFKEGVYINVYAKKGERLFINEGKILWSGFSREVFFSCSRDQ 547
QY 380 -GSTKDVSEB-----AYVCLPREG-----CPFCADD-----SPCFVQEDKYL 417
Db 548 AGTRKGIIEGPTCCFECVCEPDGEYSGETDASACDKCPDFHSNENYTSIAKEIEFLA 607
QY 418 -----LAISFOGLCMLDFVSMVYVYHFRKAKSIRASGLILLETILFGSLLLYFPVVI 471
Db 608 WTEPFGIALTFLAVLGIFLTAFLGVFIKFRNTPIVKATNRELSYLLLSLCCFSSSLF 667
QY 472 LYEPSTFCILLRWALLGCFATVYGTVLKLRVLKVPFLSRTAQRIPIYMTGGRVWRMLA 531
Db 668 FIEGPDQWTCRLRQPAFGISFVLCISILVKTNRVLVFEAKIPTSFHRKMWGLNQFL 727
QY 532 V-----ILLVFWFLIGTSSVCNLEKQISLIGOGKTSOHLINMCLIDRWDMYMTAV 584
Db 728 VFICTFMQIICIIWLYTAPPSSY-RNHELE-----DELIITC-----HEGSLMAL 773

```

Oy 585 GWMSLVSY 592
Db 774 G--SLIGY 779

RESULT 5
CASR_HUMAN STANDARD; PRT; 1078 AA.
AC P41180; Q13912; Q16379; Q16108; Q16110;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE EXTRACELLULAR CALCIUM-SENSING RECEPTOR PRECURSOR (CASR) (PARATHYROID
DE CELL CALCIUM-SENSING RECEPTOR).
GN CASR OR PCAR1.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RA Pearce S.H.S., Thakker R.V.;
RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
[2]
RA SEQUENCE FROM N.A.
RC TISSUE=Parathyroid;
RX MEDLINE=95279439; PubMed=7759551;
RA Garrett J.E., Capuano I.V., Hammerland L.G., Hung B.C., Brown E.M.,
RA Hebert S.C., Nemeth E.F., Fuller F.;
RA "Molecular cloning and functional expression of human parathyroid
RT calcium receptor cDNAs";
RL J. Biol. Chem. 270:12919-12925(1995).
[3]
RA SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=95408281; PubMed=7677761;
RA Aida K., Koishi S., Tawata M., Onaya T.;
RA "Molecular cloning of a putative Ca(2+)-sensing receptor cDNA from
RT human kidney";
RL Biochem. Biophys. Res. Commun. 214:524-529(1995).
[4]
RA SEQUENCE FROM N.A.
RX MEDLINE=96343808; PubMed=8756555;
RA Freichel M., Zink-Lorenz A., Hollloschi A., Hafner M., Flockerzi V.,
RA Raue F.;
RA "Expression of a calcium-sensing receptor in a human medullary
RT thyroid carcinoma cell line and its contribution to calcitonin
RT secretion";
RL Endocrinology 137:3842-3848(1996).
[5]
RA VARIANTS FHH GLU-185; LYS-297 AND TRP-795.
RX MEDLINE=94094324; PubMed=7916660;
RA Pollak M.R., Brown E.M., Chou Y.H., Hebert S.C., Marx S.J.,
RA Steinmann B., Levi T., Seidman C.E., Seidman J.G.;
RA "Mutations in the human Ca(2+)-sensing receptor gene cause familial
RT hypocalcemic hypercalcaemia and neonatal severe
RT hyperparathyroidism";
RL Cell 75:1297-1303(1993).
[6]
RA VARIANT ADH ALA-127.
RX MEDLINE=95179179; PubMed=7874174;
RA Pollak M.R., Brown E.M., Estep H.L., McLaine P.N., Kifor O., Park J.,
RA Hebert S.C., Seidman C.E., Seidman J.G.;
RA "Autosomal dominant hypocalcaemia caused by a Ca(2+)-sensing receptor
RT gene mutation";
RL Nat. Genet. 8:303-307(1994).
[7]
RA VARIANTS FHH MET-62; CYS-66; MET-138; GLU-143 AND GLN-227.
RX MEDLINE=95243222; PubMed=7726161;
RA Chou Y.-H.W., Pollak M.R., Brandi M.L., Toss G., Aronqvist H.,
RA Atkinson A.B., Papapoulos S.E., Marx S., Brown E.M., Seidman J.G.,
RA Seidman C.E.;
RA "Mutations in the human Ca(2+)-sensing-receptor gene that cause
RT familial hypocalcemic hypercalcaemia";
Am. J. Hum. Genet. 56:1075-1079(1995).
[8]
RA SEQUENCE OF 1-61 FROM N.A., AND VARIANT FHH ALA-39.
RX MEDLINE=95403641; PubMed=7673400;
RA Aida K., Koishi S., Inoue M., Nakazato M., Tawata M., Onaya T.;
RA "Familial hypocalcemic hypercalcaemia associated with mutation in the
RT human Ca(2+)-sensing receptor gene";
RL J. Clin. Endocrinol. Metab. 80:2594-2598(1995).
[9]
RA VARIANTS NSHPT LEU-227 AND TYR-598.
RX MEDLINE=96292293; PubMed=8675635;
RA Pearce S.H.S., Trump D., Wooding C., Besser G.M., Chew S.L.,
RA Grant D.B., Heath D.A., Hughes I.A., Paterson C.R., Whyte M.P.,
RA Thakker R.V.;
RA "Calcium-sensing receptor mutations in familial benign hypercalcaemia
RT and neonatal hyperparathyroidism";
RL J. Clin. Invest. 96:2683-2692(1995).
[10]
RA VARIANTS ADHP THR-116; HIS-681 AND SER-806, AND VARIANT SER-851.
RX MEDLINE=96311554; PubMed=8733126;
RA Baron J., Winer K.K., Yancovski J.A., Cunningham A.W., Laue L.,
RA Zimmerman D., Cutler G.B. Jr.;
RA "Mutations in the Ca(2+)-sensing receptor gene cause autosomal
RT dominant and sporadic hypoparathyroidism";
RL Hum. Mol. Genet. 5:601-606(1996).
[11]
RA VARIANT FHH ARG-174.
RX MEDLINE=97442275; PubMed=9298824;
RA Ward B.K., Stuckey B.G.A., Gutteridge D.H., Laing N.G., Pullan P.T.,
RA Ratajczak T.;
RA "A novel mutation (L174R) in the Ca2+-sensing receptor gene
RT associated with familial hypocalcemic hypercalcaemia";
RL Hum. Mutat. 10:233-235(1997).
CC -1- FUNCTION: SENSE CHANGES IN THE EXTRACELLULAR CONCENTRATION OF
CC CALCIUM IONS. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A
CC G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
CC MESSENGER SYSTEM.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS SEEM TO BE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: FOUND IN KIDNEY, BUT NOT IN BRAIN, LUNG,
CC LIVER, HEART, SKELETAL MUSCLE, OR PLACENTA.
CC -1- DISEASE: DEFECTS IN CASR ARE A CAUSE OF FAMILIAL HYPOCALCAEMIC
CC HYPERCALCAEMIA (FHH) AND NEONATAL SEVERE HYPERPARATHYROIDISM
CC (NSHPT). TWO INHERITED CONDITIONS CHARACTERIZED BY ALTERED CALCIUM
CC HOMEOSTASIS. THE MUTATIONS REDUCE THE ACTIVITY OF THE RECEPTOR.
CC FHH AFFECTED INDIVIDUALS EXHIBIT MILD OR MODEST HYPERCALCAEMIA,
CC RELATIVE HYPOCALCAEMIA, AND INAPPROPRIATELY NORMAL PTH LEVELS. IN
CC CONTRAST, NSHPT IS A RARE AUTOSOMAL RECESSIVE LIFE-THREATENING
CC DISORDER CHARACTERIZED BY VERY HIGH SERUM CALCIUM CONCENTRATIONS,
CC SKELETAL DEMINERALIZATION, AND PARATHYROID HYPERPLASIA. IN SOME
CC INSTANCES NSHPT HAS BEEN DEMONSTRATED TO BE THE HOMOZYGOUS FORM OF
CC FHH.
CC -1- DISEASE: DEFECTS IN CASR ARE ALSO THE CAUSE OF AUTOSOMAL DOMINANT
CC HYPOCALCAEMIA (ADH) IN WHICH THE RECEPTOR IS ACTIVATED AT SUBNORMAL
CC CA(2+) LEVELS.
CC -1- DISEASE: DEFECTS IN PCAR1 ARE ALSO THE CAUSE OF AUTOSOMAL DOMINANT
CC HYPOPARATHYROIDISM (ADHP). ADHP IS CHARACTERIZED BY HYPOCALCAEMIA
CC AND HYPERPHOSPHATEMIA DUE TO INADEQUATE SECRETION OF PARATHYROID
CC HORMONE. SYMPTOMS ARE SEIZURES, TETANY AND CRAMPS.
CC -1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use, by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X81086; CAA56990.1; -;
DR EMBL; U20759; AAA86503.1; -;

DR EMBL: U20760; AAA86504.1; -
 DR EMBL: D50855; BAA09453.1; -
 DR EMBL: S81176; AAB46873.1; -
 DR EMBL: S79217; AAB35262.1; -
 DR EMBL: S68032; AAB29413.2; ALT_SEQ.
 DR EMBL: S68033; AAB29414.1; -
 DR EMBL: S68036; AAB29415.1; -
 DR GCRDB: GCR_1337; -
 DR GCRDB: GCR_1874; -
 DR GCRDB: GCR_2012; -
 DR GCRDB: GCR_2013; -
 DR GCRDB: GCR_2696; -
 DR GCRDB: GCR_2697; -
 DR MIM: 601199; -
 DR MIM: 145980; -
 DR MIM: 601198; -
 DR InterPro: IPR000068; -
 DR InterPro: IPR000337; -
 DR InterPro: IPR001828; -
 DR Pfam: PF00003; 7tm_3; 1.
 DR PRINTS: PR00248; GPCRMR.
 DR PRINTS: PR00592; CASENSINGR.
 DR PROSITE: PS00879; G_PROTEIN_RECEP_F3_1; 1.
 DR PROSITE: PS00980; G_PROTEIN_RECEP_F3_2; 1.
 DR PROSITE: PS00981; G_PROTEIN_RECEP_F3_3; 1.
 DR PROSITE: PS00259; G_PROTEIN_RECEP_F3_4; 1.
 DR G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
 KW Disease mutation; Alternative splicing; Polymorphism.
 FT SIGNAL 1 19
 FT CHAIN 20 1078
 FT DOMAIN 20 612
 FT TRANSMEM 613 635
 FT DOMAIN 636 649
 FT TRANSMEM 650 670
 FT DOMAIN 671 681
 FT TRANSMEM 682 700
 FT DOMAIN 701 724
 FT TRANSMEM 725 745
 FT DOMAIN 746 769
 FT TRANSMEM 770 792
 FT DOMAIN 793 805
 FT TRANSMEM 806 828
 FT DOMAIN 829 836
 FT TRANSMEM 837 878
 FT DOMAIN 863 1078
 FT CARBOHYD 90 90
 FT CARBOHYD 130 130
 FT CARBOHYD 261 261
 FT CARBOHYD 287 287
 FT CARBOHYD 386 386
 FT CARBOHYD 400 400
 FT CARBOHYD 446 446
 FT CARBOHYD 468 468
 FT CARBOHYD 488 488
 FT CARBOHYD 541 541
 FT CARBOHYD 594 594
 FT VARSPLIC 536 536
 FT VARIANT 39 39
 FT VARIANT 52 52
 FT VARIANT 62 62
 Query Match 4.0%; Score 127; DB 1; Length 1078;
 Best Local Similarity 22.5%; Pred. No. 0.045;
 Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;
 QY 311 QKVDIDQSSDGFSTHCK---HLNNS-CMPKGLGVLCAYECICAG---FVHPGV 363
 DB 475 EQVTFDEC---GDLVGNYSIIHNLSPEDGSIVFKEVG---YNNYAKKGERLFTNEEK 527
 QY 364 LPVNNRRRGPDQH-----ISGSTKDVSSE---AVVCLPDCREG-----CPFCAD 404
 DB 528 ILWSGFSRVPFNSRDLCTAGTRKGIIEGPTCCFCECPDGEYSDTDSACKNCPD 587

QY 405 D-----SPCFVQEDKYLR-----LAIISFQGLCMLLDVFSMLVYVYHFRKAKSIRASG 451
 DB 588 DPWSNENHTSCIATKEITFLSWTEPFGIAULTFAVLGIELTAFLGVFIFKFRNTPIVKATN 647
 QY 452 LILLETILFGSLLLYFPVIVILPEPSTFRCCILLRWALLGFGATVYGTVTLKLRVLRVFL 511
 DB 648 RELSYLLFLSLCCFSFSSFFIGEPQDWTCRLRQPAFGISFVLCISCLVKNRVLVFE 707
 QY 512 SRTAQRIPTYMTGGRVMRLAV-----ILLVVFELIGWTSSV-CONLEKOISLI----- 559
 DB 708 AKIPTSFHRKWMGLNLQFLVFLCTFMQIVICVIMLYTAPPSSYRNQOELEDEIFITCHE 767
 QY 560 GOGKTSDDLIFNMCLI 575
 DB 768 GSLMALGFLIGYTCLL 783
 RESULT 6
 CASR_BOVIN STANDARD; PRT: 1085 AA.
 AC P35384;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE EXTRACELLULAR CALCIUM-SENSING RECEPTOR PRECURSOR (CASR) (PARATHYROID
 DE CELL CALCIUM-SENSING RECEPTOR).
 GN CASR OR PCARL.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 RN NCBI_TaxID=9913;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Parathyroid;
 RA MEDLINE=94077182; PubMed=8255296;
 RA Brown E.M., Gamba G., Riccardi D., Lombardi M., Butters R., Kifor O.,
 RA Sun A., Hediger M.A., Lytton J., Hebert S.C.;
 RT "Cloning and characterization of an extracellular Ca(2+)-sensing
 RT receptor from bovine parathyroid.";
 RL Nature 366:575-580(1993).
 CC -!- FUNCTION: SENSE CHANGES IN THE EXTRACELLULAR CONCENTRATION OF
 CC CALCIUM IONS. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A
 CC G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
 CC MESSENGER SYSTEM.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: S67307; AAB29171.1; -
 DR PIR: S40476; S40476.
 DR GCRDB: GCR_0900; -
 DR InterPro: IPR000068; -
 DR InterPro: IPR000337; -
 DR InterPro: IPR001828; -
 DR Pfam: PF00003; 7tm_3; 1.
 DR Pfam: PF01094; ANF_receptor; 1.
 DR PRINTS: PR00248; GPCRMR.
 DR PRINTS: PR00592; CASENSINGR.
 DR PROSITE: PS00979; G_PROTEIN_RECEP_F3_1; 1.
 DR PROSITE: PS00980; G_PROTEIN_RECEP_F3_2; 1.
 DR PROSITE: PS00981; G_PROTEIN_RECEP_F3_3; 1.
 DR PROSITE: PS00259; G_PROTEIN_RECEP_F3_4; 1.
 DR G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.
 KW SIGNAL 1 19
 FT


```

Db 520 WLCIPQ---PYEYRDEFTCADCGLYWPNASLTGCFELPQEIYIRGDAWAVGPVTIAC 576
QY 423 FQGLCMLLDVSMVYVYHFRKAKSIRASGLILLETILFGSLILYFVWVILYXPPSTFRCI 482
Db 577 LGALATL--FVLGVFVRH-NATPVKASGRELVCYILGGVFLCYCMTFEIAKPSSTAVCT 633
QY 483 LLRWALLGFATVYGTVTLKLRVLFV--LSRTAQIRPYMT-GGRVWMLAVI-----LL 535
Db 634 LRLGLGTAFSVCSYALLTKTNRIARIFGGAREGAORPRFISPASQVAICLALISGOLLI 693
QY 536 VVWFVFLI 542
Db 694 VVAVLVV 700

RESULT 8
GR2_RAT GR2_RAT STANDARD; PRT; 872 AA.
AC MGR2_RAT P31421;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 2 PRECURSOR.
GN GRM2 OR MGLUR2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92110002; PubMed=1309649;
RA Tanabe Y., Masu M., Ishii T., Shigemoto R., Nakanishi S.;
RT "A family of metabotropic glutamate receptors.";
RL Neuron 8:169-179(1992).
CC -!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS
CC MEDIATED BY A G-PROTEIN THAT INHIBITS ADENYLATE CYCLASE ACTIVITY.
CC MAY MEDiate SUPPRESSION OF NEUROTRANSMISSION OR MAY BE INVOLVED IN
CC SYNAPTOGENESIS OR SYNAPTIC STABILIZATION.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: IS WIDELY DISTRIBUTED IN THE CNS AND PROMINENT
CC EXPRESSION IS SEEN IN GOLGI CELLS OF THE CEREBELLUM AND SOME
CC PARTICULAR NEURONAL CELLS IN OTHER BRAIN REGIONS.
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC STRONGEST, TO MGLUR3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M92075; ; NOT_ANNOTATED_CDS.
CC PIR; JH0561; JH0561.
CC DR HSSP; P06612; 1ECL.
CC DR GCRdb; GCR_0361; -.
CC DR InterPro; IPR000162; -.
CC DR InterPro; IPR000337; -.
CC DR InterPro; IPR001458; -.
CC DR InterPro; IPR001828; -.
CC DR Pfam; PF00003; 7tm_3; 1.
CC DR PRINTS; PF01094; ANF_receptor; 1.
CC DR PRINTS; PR00248; GPCRMR.
CC DR PRINTS; PR00593; MTABOTROPICR.
CC DR PRINTS; PR01052; MTABOTROPICR.
CC DR PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1.
CC DR PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1.
CC DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
CC DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
CC DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_4; 1.
CC DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_4; 1.
CC DR G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;

```

```

KW Multigene family. 18
FT SIGNAL 19 872
FT CHAIN 19 567
FT DOMAIN 19 567
FT TRANSMEM 568 590
FT DOMAIN 591 604
FT TRANSMEM 605 625
FT DOMAIN 626 636
FT TRANSMEM 637 655
FT DOMAIN 656 679
FT TRANSMEM 680 700
FT DOMAIN 701 725
FT TRANSMEM 726 747
FT DOMAIN 748 760
FT TRANSMEM 761 783
FT DOMAIN 784 793
FT TRANSMEM 794 819
FT DOMAIN 820 872
FT CARBOHYD 203 203
FT CARBOHYD 286 286
FT CARBOHYD 338 338
FT CARBOHYD 402 402
FT CARBOHYD 547 547
SQ SEQUENCE 872 AA; 95773 MW; 1E74CABD6AD4BED9 CRC64;

Query Match 3.6%; Score 116; DB 1; Length 872;
Best Local Similarity 18.9%; Pred No. 0.25;
Matches 93; Conservative 63; Mismatches 184; Indels 152; Gaps 22;

QY 141 QSNKRE-----QNLQDLDWYQALVWSLLE----GEPSISRAAITFTSDLSAPAPQVF 191
Db 271 RSEDARELLAATQRLNASFTWASDQWGALESVAGSERAEGATIELASYPISDFASY 330
QY 192 LQATREESRILLQDLSSSAPHLANATLETWFHGLRRKWRPHLRRGPNQGRGLGHSWR 251
Db 331 FQS-----LDPWNS-----RNPWF---REFWEERFH-----CSFR 358
QY 252 RKDLGGDKSHKSPYPLECENGSKYKGMVLTLSAIVGLQPNLVP----- 298
Db 359 ORD-----CAHSLRAVPEQESKIMFVNVNAVYAMAHALNHMRALCPNTHLCDAMRPVN 414
QY 299 -----EFGVMKVDINLQKVDI-DQSSDGMFSGTHKCHLNSECMPKGLGFLVGLAYE 351
Db 415 GRRLYKDFVNLVKFADPAPRPAATDDEVRDFREGDG-----IGRYN 454
QY 352 CIC-----KAGFYHPGVLPVNNF-----RRRCPDQHSISTKDYSE 387
Db 455 IFTYLRAGSGRYRYOKYGVYWAEGSLTDSFIPWASPSAGPLPASRCSEPCLNQKVSQVP 514
QY 388 E---AYVCLPCREGCPF-----CADD-----SPCFVQEDKYLR----- 417
Db 515 GEVCCWLCIPQ---PYEYRDEFTCADCGLYWPNASLTGCFELPQEIYIRGDAWAVGP 571
QY 418 LAIISFQGLCMLLDVSMVYVYHFRKAKSIRASGLILLETILFGSLILYFVWVILYXPP 477
Db 572 VTIAIGALATL--FVLGVFVRH-NATPVKASGRELVCYILGGVFLCYCMTFEIAKPS 628
QY 478 TPRCILLRWALLGFATVYGTVTLKLRVLFV--LSRTAQIRPYMT-GGRVWMLAVI- 533
Db 629 TAVCTLRRLGLGTAFSVCSYALLTKTNRIARIFGGAREGAORPRFISPASQVAICLALIS 688
QY 534 ---LLVWFVFLI 542
Db 689 GOLLIVAAWLVV 700

RESULT 9
MGR1_HUMAN STANDARD; PRT; 1194 AA.
AC O13255; O13256; O14757; O14758;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

```

01-OCT-2000 (Rel. 40, Last annotation update)
 METABOTROPIC GLUTAMATE RECEPTOR 1 PRECURSOR.
 GRI1 OR MGLUR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96029774; PubMed=7476890;
 RA Desai M.A., Burnett J.P., Mayne N.G., Schoepp D.D.;
 RT "Cloning and expression of a human metabotropic glutamate receptor 1
 alpha: enhanced coupling on co-transfection with a glutamate
 transporter.";
 RT Mol. Pharmacol. 48:648-657(1995).
 RL [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=97231349; PubMed=9076744;
 RA Stephan D., Bon C., Holzwarth J.A., Galvan M., Pruss R.M.;
 RT "Human metabotropic glutamate receptor 1: mRNA distribution,
 RT chromosome localization and functional expression of two splice
 RT variants.";
 RT Neuropharmacology 35:1649-1660(1996).
 RL [3]
 CC -!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS
 CC MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-
 CC CALCIUM SECOND MESSENGER SYSTEM. MAY PARTICIPATE IN THE CENTRAL
 CC ACTION OF GLUTAMATE IN THE CNS, SUCH AS LONG-TERM POTENTIATION IN
 CC THE HIPPOCAMPUS AND LONG-TERM DEPRESSION IN THE CEREBELLUM.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
 CC STRONGEST, TO MGLUR5.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U31215; AAA87843.1; -;
 CC EMBL; U31216; AAA87844.1; -;
 CC EMBL; L76627; AAB05337.1; -;
 CC EMBL; L76631; AAB05338.1; -;
 CC MIM; 604473; -;
 CC GCRDb; GCR_1825; -;
 CC GCRDb; GCR_1826; -;
 CC GCRDb; GCR_1982; -;
 CC GCRDb; GCR_1983; -;
 CC InterPro; IPR000162; -;
 CC InterPro; IPR000337; -;
 CC InterPro; IPR001256; -;
 CC InterPro; IPR001828; -;
 CC Pfam; PF00003; 7tm_3; 1.
 CC Pfam; PF01094; ANF_receptor; 1.
 CC PRINTS; PR00248; GPCRMR.
 CC PRINTS; PR00593; METABOTROPIC1R.
 CC PRINTS; PR01051; METABOTROPIC1R.
 CC PROSITE; PS00979; G_PROTEIN_RECEPTOR_F3_1; 1.
 CC PROSITE; PS00980; G_PROTEIN_RECEPTOR_F3_2; 1.
 CC PROSITE; PS00981; G_PROTEIN_RECEPTOR_F3_3; 1.
 CC PROSITE; PS02029; G_PROTEIN_RECEPTOR_F3_4; 1.
 CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
 CC Multigene family; Alternative splicing.
 CC SIGNAL 1 18 POTENTIAL.
 CC CHAIN 19 1194 METABOTROPIC GLUTAMATE RECEPTOR 1.
 CC DOMAIN 19 592 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 593 615 I (POTENTIAL).
 CC DOMAIN 616 629 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 630 650 II (POTENTIAL).
 CC DOMAIN 651 661 EXTRACELLULAR (POTENTIAL).
 CC

FT TRANSMEM 662 680
 FT DOMAIN 681 706
 FT TRANSMEM 707 727
 FT DOMAIN 728 750
 FT TRANSMEM 751 772
 FT DOMAIN 773 785
 FT TRANSMEM 786 808
 FT DOMAIN 809 814
 FT TRANSMEM 815 840
 FT DOMAIN 841 1194
 FT DOMAIN 1014 1035
 FT DOMAIN 1067 1081
 FT DOMAIN 1095 1130
 FT DOMAIN 1142 1194
 FT CARBOHYD 98 98
 FT CARBOHYD 223 223
 FT CARBOHYD 397 397
 FT CARBOHYD 515 515
 FT VARSPLIC 887 906
 FT VARSPLIC 907 1194
 FT VARSPLIC 887 906
 FT CONFLICT 593 593
 FT SEQUENCE 1194 AA; 132376 MW; 970E51AF40584F40 CRC64;
 Query Match 3.6%; Score 116; DB 1; Length 1194;
 Best Local Similarity 22.1%; Pred. No. 0.38;
 Matches 49; Conservative 34; Mismatches 95; Indels 44; Gaps 6;
 QY 357 GFYHPGVLPVNNRRRGGPDQHSIGSTKDVSEE-----AVVCLPCRGG- 398
 Db 498 GTWHEGVNIDDDYKIQ---MKNKGVRSVCSEPCLGKQIKVIRKEVSCCCTACKENE 554
 QY 399 -----CPCF-----ADSPCFVQEDKYLR-----LAIISFGICMLLDVFMVLV 438
 Db 555 YVQDEFTCKACDLGWPNNADLTGCEPIPVRYLEWSNIEPIIAIAFSCILGILVTLFVTLIF 614
 QY 439 YHFKAKSIRASGLITLLETILFGSLLYFPVVLVYFPEPTEPCILLRWLRLLGATVYGT 498
 Db 615 VLYRTPVWKSRSRELCYIILAGIFLGVCFPTLIARPTTSCYLLRLLVGLSAMCYSA 674
 QY 499 VTLKHLRVKVL---SRTAQRIPYMTGGRVNRMLAVILLV 537
 Db 675 LVTKNRIARILAGSKKIKCTKPRFMSAWAQVIAISLISV 716
 RESULT 10
 MGR3_HUMAN STANDARD; PRT; 877 AA.
 ID MGR3_HUMAN
 AC Q14832;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE METABOTROPIC GLUTAMATE RECEPTOR 3 PRECURSOR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96437205; PubMed=8840013;
 RA Makoff A., Volpe F., Lechuk R., Harrington K., Emson P.;
 RT "Molecular characterization and localization of human metabotropic
 RT glutamate receptor type 3.";
 RL Brain Res. Mol. Brain Res. 40:55-63(1996).
 CC -!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR
 CC IS MEDIATED BY A G-PROTEIN THAT INHIBITS ADENYLATE CYCLASE
 CC ACTIVITY.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC STRONGEST, TO MGLUR2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X77748; CAA54796.1; -
CC GCRDb: GCR_2070; -
CC MIM: 601115; -
CC InterPro: IPR000162; -
CC InterPro: IPR000337; -
CC InterPro: IPR001234; -
CC InterPro: IPR001828; -
CC Pfam: PF00003; 7tm_3; 1.
CC Pfam: PF01094; ANF_receptor; 1.
CC PRINTS: PR00248; GPCRMR.
CC PRINTS: PR00593; MTABOTROPICR.
CC PRINTS: PR01053; MTABOTROPICR.
CC PROSITE: PS00979; G_PROTEIN_RECEP_F3_1; 1.
CC PROSITE: PS00980; G_PROTEIN_RECEP_F3_2; 1.
CC PROSITE: PS00981; G_PROTEIN_RECEP_F3_3; 1.
CC PROSITE: PS0259; G_PROTEIN_RECEP_F3_4; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
CC Multigene family.
CC SIGNAL 1 20
CC CHAIN 21 877
CC DOMAIN 21 574
CC TRANSMEM 575 597
CC DOMAIN 598 611
CC TRANSMEM 612 632
CC DOMAIN 633 643
CC TRANSMEM 644 662
CC DOMAIN 663 686
CC TRANSMEM 687 707
CC DOMAIN 708 732
CC TRANSMEM 733 754
CC DOMAIN 755 767
CC TRANSMEM 768 790
CC DOMAIN 791 800
CC TRANSMEM 801 826
CC DOMAIN 827 877
CC CARBOHYD 207 207
CC CARBOHYD 290 290
CC CARBOHYD 412 412
CC CARBOHYD 437 437
CC SEQUENCE 877 AA; 98619 MW; 66F28663CE35F740 CRC64;
Query Match 3.6%; Score 115.5; DB 1; Length 877;
Best Local Similarity 18.6%; Pred. No. 0.27;
Matches 90; Conservative 72; Mismatches 196; Indels 125; Gaps 23;
QY 138 VMLQSKSRE-----ONLQDDLDWQALVW-----SLLEGPSISRAAIFSTDSLAPAP 188
DB 272 LFMRSDDSLRIAASRANASTFWASDCGAGOEIISKSEHVAYCAITLLEAS----- 325
QY 189 QVFLQATRESRIILQDLSSAPHLANATLETWFHGLRRKRWPHLHRRGPNQGRGLGH 248
DB 326 ----QPVROFDR-YFQSLNPYNNH-----RNPWFRDE----- 352
QY 249 SWRRK--DGLGDKSHFKWSPPLCEGNSYKP-----GWLVTLSAAYGLOPNLVP 298
DB 353 -WEQFQCSLQKRNHRRVCDKHLAIDSSNYQESKIMFVNAVAVAMALHMKMTLCP 411
QY 299 ----EPRGVKMDINLQKVDICQSSDGFSGTHKCHLNHSECMPIKGLGFLVGLAYECI-- 353
DB 412 NTKLCDAWKI-LDGKLYKDYLLKIN-FTAPFPNPKDADSVKFTDGDGNGRYNVNF 469

QY 354 -----CKAGFYHPGV-LPVNPF---RRGPDQHIS-----GSTKDVSEB---AYVC 392
DB 470 QNVGGKYSYLVKGHWAETLSLDVNSHSNSVPTSCSDPCAPNEMKNMQPDGVCCWIC 529
QY 393 LPCR-----GCPFC-----ADDSPCFVQEDKYLK-----LAIISFOGL 426
DB 530 IPECEVEYLADEFTCMDCGQWPTADLTGCDYLPEDYIRWEDAWAIGPVTTIACLGFMCT 589
QY 427 CMLLDFVSMVYVHFRAKASIRASGILLETILFGLSLLYFPVVIYLFEPSTFCILLRW 486
DB 590 CMV-----VTVEIKHNNTPLVRASGRELVCYILLFGVGLSCYMTFFFTAKPSPICALRL 644
QY 487 ARLLGPAFTVGTTLKLRVLFV---LSRTAQRIPTWT---GGRVMRLAVTL-----LVFV 539
DB 645 GLGSFAICYALLTKTNCIARIFDGVKNGAQRKFTSPSSQVFICLGLILVQIVMSVW 704
QY 540 FLI 542
DB 705 LIL 707
RESULT 11
MGLR1_RAT
ID MGLR1_RAT STANDARD; PRT; 1199 AA.
AC P23385;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 1 PRECURSOR.
GN GRM1 OR MGLUR1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX MEDLINE=91156047; PubMed=1847995;
RA Masu M., Tanabe Y., Tsuchida K., Shigemoto R., Nakanishi S.;
RT "Sequence and expression of a metabotropic glutamate receptor.";
RL Nature 349:760-765(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92022526; PubMed=1656524;
RA Houamed K.M., Kujper J.L., Gilbert T.L., Haldeman B.A., O'Hara P.J.,
RA Mulvihill E.R., Almers W., Hagen F.S.;
RT "Cloning, expression, and gene structure of a G protein-coupled
RT glutamate receptor from rat brain.";
RL Science 252:1318-1321(1991).
RN [3]
RP ALTERNATIVE SPLICING (ISOFORM 1B).
RC TISSUE=Brain;
RX MEDLINE=92110002; PubMed=1309649;
RA Tanabe Y., Masu M., Ishii T., Shigemoto R., Nakanishi S.;
RT "A family of metabotropic glutamate receptors.";
RL Neuron 8:169-179(1992).
RN [4]
RP ALTERNATIVE SPLICING (ISOFORM 1C).
RC TISSUE=Brain;
RX MEDLINE=93066232; PubMed=1438218;
RA Pin J.-P., Waerber C., Prezeau L., Bockaert J., Heinemann S.F.;
RT "Alternative splicing generates metabotropic glutamate receptors
RT inducing different patterns of calcium release in Xenopus oocytes.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:10331-10335(1992).
CC -!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS
CC MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-
CC CALCIUM SECOND MESSENGER SYSTEM. MAY PARTICIPATE IN THE CENTRAL
CC ACTION OF GLUTAMATE IN THE CNS, SUCH AS LONG-TERM DEPRESSION IN
CC THE HIPPOCAMPUS AND LONG-TERM DEPRESSION IN THE CEREBELLUM.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1A (SHOWN HERE), 1B AND 1C;
CC ARE PRODUCED BY ALTERNATIVE SPLICING. 1B AND 1C ARE C-TERMINALLY

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC EMBL; D28538; BAA05891.1; -
CC EMBL; D28539; BAA05892.1; -
CC EMBL; S64316; AAD13954.1; -
DR GCRDB; GCR_0761; -
DR GCRDB; GCR_1002; -
DR GCRDB; GCR_1003; -
DR GCRDB; GCR_1317; -
DR MIM; 604102; -
DR InterPro; IPR000162; -
DR InterPro; IPR000202; -
DR InterPro; IPR000337; -
DR InterPro; IPR001828; -
DR Pfam; PF00003; 7tm_3; 1.
DR PRINTS; PR00248; GPCRMR.
DR PRINTS; PR00593; MTABOTROPICR.
DR PRINTS; PR01085; MTABOTROPICR.
DR PROSITE; PS00979; G_PROTEIN_RECF_F3_1; 1.
DR PROSITE; PS00980; G_PROTEIN_RECF_F3_2; 1.
DR PROSITE; PS00981; G_PROTEIN_RECF_F3_3; 1.
DR PROSITE; PS00982; G_PROTEIN_RECF_F3_4; 1.
DR G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Multigene family; Alternative splicing.
FT SIGNAL 1 20
FT CHAIN 21 1212
FT DOMAIN 22 579
FT TRANSMEM 580 602
FT DOMAIN 603 616
FT TRANSMEM 617 637
FT DOMAIN 638 648
FT TRANSMEM 649 667
FT DOMAIN 668 693
FT TRANSMEM 694 714
FT DOMAIN 715 737
FT TRANSMEM 738 759
FT DOMAIN 760 772
FT TRANSMEM 773 795
FT DOMAIN 796 801
FT TRANSMEM 802 827
FT DOMAIN 828 1212
FT CARBOHYD 88
FT CARBOHYD 210
FT CARBOHYD 378
FT CARBOHYD 382
FT CARBOHYD 445
FT CARBOHYD 445
FT CARBOHYD 734
FT CARBOHYD 773
FT VARSPLIC 877
SQ SEQUENCE 1212 AA; 132468 MW; A3C73606681C6A25 CRC64;

Query Match
Best Local Similarity 3.6%; Score 114; DB 1; Length 1212;
Matches 80; Conservative 53; Mismatches 165; Indels 94; Gaps 16;

QY 219 ETEWFLHRRKWRPHLHRRGNQGRPLGHSWR-----RKDGLGGDKSHFKWSPPLYECE 273
DB 333 DVKWFDDYLLKRPETNHRNP-----WFQEFQHRFQCRLEGFPQENSKYNT-----CN 382

QY 274 NG-----SYKPGWLW-TLSSAIYGL---QPNLVPEFRGV-----MKVDINLQ 311
DB 383 SSLTKLTHVQSDKMGFVFNATYSMAYGHLNMQMSLCPGYAGLCDAKPIDGRKLLSLM 442

QY 312 KYVIDQCSGDFWFGTHKHLNNSCMPKGLGLGVGLGAYECICKAGFYHPGVLPLVNNFR 371
DB 443 KTNFTGVSGDTLFDENGDSRGRYEIMNFKMG-----KDYFDIYNGVSDNGEL 492

QY 372 RGPDOHI-----SGSTKDVSEE-----AYVCLPCREG-----CPPC 402
DB 493 KMDDEVMSKSNIIIRSVSCPECKGQIKVIRKGEVSCWCTCTPCKENEYVFDEYTCAC 552
```

```
QY 403 -----ADD-SPCFVQEDKYLR-----LAIISFQGLCMLLDFVSMVYVHFRKAKSIRA 449
DB 553 QUGSWPTDITLGTGDLIPVQYLRMGDPEPTAAVAVFACLGGLATLFTVTVFIYRDTVPVK 612
QY 450 SGLILLETILFGLSLLYFPVILYFSPFRICLLRWARLIGFATVYGTVTLKLHRLVKV 509
DB 613 SSRELCTIILAGICGLYLCFTCLIAKPKQIYCYLQIGIGLSFAMSYSAVVTNTRIARI 672
QY 510 FLSRTAQRI-----PYMTGGRVMRMLAVILLV 537
DB 673 -LAGSKKICKTKKPREMSACAQLVIAFILICI 703

RESULT 13
MGR3_RAT
ID MGR3_RAT STANDARD; PRT; 879 AA.
AC P31422;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 3 PRECURSOR.
GN GRM3 OR MGLUR3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92110002; PubMed=1309649;
RA Tanabe Y., Masu M., Ishii T., Shigemoto R., Nakanishi S.;
RT "A family of metabotropic glutamate receptors.";
RL Neuron 8:169-179(1992).
CC !- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR
CC IS MEDIATED BY A G-PROTEIN THAT INHIBITS ADENYLATE CYCLASE
CC ACTIVITY.
CC !- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC !- TISSUE SPECIFICITY: IS WIDELY DISTRIBUTED IN THE CNS. PREDOMINANT
CC EXPRESSION IS SEEN IN THE NEURONAL CELLS OF THE CEREBRAL CORTEX,
CC DENTATE GYRUS, AND GLIAL CELLS THROUGHOUT BRAIN REGIONS.
CC !- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC STRONGEST, TO MGLUR2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

CC EMBL; M92076; -; NOT_ANNOTATED_CDS.

DR PIR; JH0562; JH0562.

DR GCRDB; GCR_03624.

DR InterPro; IPR000162; -.

DR InterPro; IPR000337; -.

DR InterPro; IPR001234; -.

DR InterPro; IPR001828; -.

DR Pfam; PF00003; 7tm_3; 1.

DR Pfam; PF01094; ANF_receptor; 1.

DR PRINTS; PR00248; GPCRMR.

DR PRINTS; PR00593; MTABOTROPICR.

DR PRINTS; PR01053; MTABOTROPICR.

DR PROSITE; PS00979; G_PROTEIN_RECF_F3_1; 1.

DR PROSITE; PS00980; G_PROTEIN_RECF_F3_2; 1.

DR PROSITE; PS00981; G_PROTEIN_RECF_F3_3; 1.

DR PROSITE; PS00259; G_PROTEIN_RECF_F3_4; 1.

DR G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;

KW Multigene family.

FT SIGNAL 1 22

FT CHAIN 23 879

FT DOMAIN 23 576

FT TRANSMEM 577 599

POTENTIAL.

METABOTROPIC GLUTAMATE RECEPTOR 3.

EXTRACELLULAR (POTENTIAL).

I (POTENTIAL).

```
FT DOMAIN 600 613 CYTOPLASMIC (POTENTIAL).
FT TRANSLEM 614 634 II (POTENTIAL).
FT DOMAIN 635 645 EXTRACELLULAR (POTENTIAL).
FT TRANSLEM 646 664 III (POTENTIAL).
FT DOMAIN 665 688 CYTOPLASMIC (POTENTIAL).
FT TRANSLEM 689 709 IV (POTENTIAL).
FT DOMAIN 710 734 EXTRACELLULAR (POTENTIAL).
FT TRANSLEM 735 756 V (POTENTIAL).
FT DOMAIN 757 769 CYTOPLASMIC (POTENTIAL).
FT TRANSLEM 770 792 VI (POTENTIAL).
FT DOMAIN 793 802 EXTRACELLULAR (POTENTIAL).
FT TRANSLEM 803 828 VII (POTENTIAL).
FT DOMAIN 829 879 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 439 439 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 879 AA; 98959 MW; 3E5955EDD5E6DEED CRC64;

Query Match 3.5%; Score 112; DB 1; Length 879;
Best Local Similarity 22.6%; Pred. No. 0.52;
Matches 42; Conservative 30; Mismatches 76; Indels 38; Gaps 7;

Qy 390 YVCLPCRE-----GCPFC-----ADSPCFVQEDKYLK-----LAISF 423
Db 529 WICIPCPEYELVDFTCMDCGPGQWPTADLSGCYNLPEDYIKWEDAWAIGVPTIACIGF 588
Qy 424 QGLCMLDFVSMVYHFRKAKSIRASGLILLETILFGLSLLYFPVWILYEPSTFCIL 483
Db 589 LCTCIVI-----TVPIKHNTPLVKASGRELCLYLLFGVLSYCNWTFEFAKPSVPICAL 643
Qy 484 LRMALLGFATVGVPTLKLHRLVKF--LSRTAQIPYMT--GGVVRMLAVIL----LV 536
Db 644 RRLGLGTSFAICYSALLTKTNCIARIFDGVKNAGRPKRFISPSQVFCILGLILVQIVMV 703
Qy 537 VFWFLI 542
Db 704 SVWLIL 709

RESULT 14
MGR5_RAT STANDARD; PRT; 1203 AA.
ID MGR5_RAT
AC P31424;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 5 PRECURSOR.
GN GRM5 OR MGLUR5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92317054; PubMed=1320017;
RA Abe T., Sugihara H., Nawa H., Shigemoto R., Mizuno N., Nakanishi S.;
RT "Molecular characterization of a novel metabotropic glutamate
RT receptor mGLUR5 coupled to inositol phosphate/Ca2+ signal
RT transduction."
RL J. Biol. Chem. 267:13361-13368(1992).
RN [2]
RP SEQUENCE OF 859-923 FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain;
RX MEDLINE=93343913; PubMed=7688218;
RA Minakami R., Katsuki F., Sugiyama H.;
RT "A variant of metabotropic glutamate receptor subtype 5: an
RT evolutionally conserved insertion with no termination codon."
RL Biochem. Biophys. Res. Commun. 194:622-627(1993).
CC -1- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS
CC MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-
```

```
CC CALCIIUM SECOND MESSENGER SYSTEM AND GENERATES A CALCIUM-ACTIVATED
CC CHLORIDE CURRENT.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 5A (SHOWN HERE) AND 5B; ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER BY AN INSERTION OF
CC 32 RESIDUES
CC -1- TISSUE SPECIFICITY: WIDELY DISTRIBUTED IN NEURONAL CELLS OF THE
CC CENTRAL NERVOUS SYSTEM.
CC -1- MISCELLANEOUS: ACTIVATED BY QUISQUALATE > GLUTAMATE > IBOTENATE >
CC TRANS-1- AMINOCYCLOPENTYL-1,3-DICARBOXYLATE.
CC -1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC STRONGEST, TO MGLUR1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D10891; -; NOT_ANNOTATED_CDS.
CC DR EMBL; S64315; AAB27666.1; -;
CC DR PIR; A42916; A42916.
CC DR GCRdb; GCR_0444; -;
CC DR GCRdb; GCR_0760; -;
CC DR InterPro; IPR000162; -;
CC DR InterPro; IPR000202; -;
CC DR InterPro; IPR000337; -;
CC DR InterPro; IPR001828; -;
CC DR Pfam; PF00003; 7cm3; 1.
CC DR Pfam; PF01094; ANF_receptor; 1.
CC DR PRINTS; PR00248; GPCRMR.
CC DR PRINTS; PR00593; MTABOTROPICR.
CC DR PRINTS; PR01055; MTABOTROPIC5R.
CC DR PROSITE; PS00979; G_PROTEIN_RECP_F3_1; 1.
CC DR PROSITE; PS00980; G_PROTEIN_RECP_F3_2; 1.
CC DR PROSITE; PS00981; G_PROTEIN_RECP_F3_3; 1.
CC DR PROSITE; PS0259; G_PROTEIN_RECP_F3_4; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
CC Multigene family; Alternative splicing.
CC SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1203 METABOTROPIC GLUTAMATE RECEPTOR 5.
FT DOMAIN 22 578 EXTRACELLULAR (POTENTIAL).
FT TRANSLEM 579 601 I (POTENTIAL).
FT DOMAIN 602 615 CYTOPLASMIC (POTENTIAL).
FT TRANSLEM 616 636 II (POTENTIAL).
FT DOMAIN 637 647 EXTRACELLULAR (POTENTIAL).
FT TRANSLEM 648 666 III (POTENTIAL).
FT DOMAIN 667 692 CYTOPLASMIC (POTENTIAL).
FT TRANSLEM 693 713 IV (POTENTIAL).
FT DOMAIN 714 736 EXTRACELLULAR (POTENTIAL).
FT TRANSLEM 737 758 V (POTENTIAL).
FT DOMAIN 759 771 CYTOPLASMIC (POTENTIAL).
FT TRANSLEM 772 794 VI (POTENTIAL).
FT DOMAIN 795 800 EXTRACELLULAR (POTENTIAL).
FT TRANSLEM 801 826 VII (POTENTIAL).
FT DOMAIN 827 1203 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 377 377 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 733 733 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 876 907 MISSING (IN ISOFORM 5A).
SQ SEQUENCE 1203 AA; 131885 MW; 99CA51E9E1ICIEA4 CRC64;
```

```
Query Match 3.5%; Score 111; DB 1; Length 1203;
Best Local Similarity 20.8%; Pred. No. 0.95;
Matches 80; Conservative 54; Mismatches 171; Indels 80; Gaps 17;

Qy 219 ETEWPHGLRRKWRPHLRHRRGNQGRGLGHSWR-----RKDGLGDKSHFKWSPYLECE 273
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 10:19:21 ; Search time 61.87 Seconds
(without alignments)
1280.923 Million cell updates/sec

Title: US-09-775-181-4
Perfect score: 3199
Sequence: 1 MGNMAYPLLLCLLLLAQLGLG.....YMTAVGMWSLVSDGLTIFQ 599

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL16: *
1: sp_Archaea: *
2: sp_Bacteria: *
3: sp_Fungi: *
4: sp_Human: *
5: sp_Invertebrate: *
6: sp_Mammal: *
7: sp_MHC: *
8: sp_Organelle: *
9: sp_Phage: *
10: sp_Plant: *
11: sp_Rodent: *
12: sp_Unclassified: *
13: sp_Vertebrate: *
14: sp_Virus: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	339	10.6	669	5 Q9VR40	Q9vr40 drosophila
2	238.5	7.5	176	5 Q9WIP4	Q9wip4 drosophila
3	161.5	5.0	976	5 Q9V485	Q9v485 drosophila
4	149	4.7	1677	5 Q9VKA3	Q9vka3 drosophila
5	132.5	4.1	868	13 Q9V3636	Q9v3636 fugu rubrip
6	130	4.1	1305	5 Q9VPS7	Q9vps7 drosophila
7	127.5	4.0	840	11 Q9Z0R8	Q9z0r8 rattus norv
8	125	3.9	977	13 Q9PWE1	Q9pwe1 ictalurus p
9	124.5	3.9	528	5 Q9G954	Q9g954 geodia cydo
10	120.5	3.8	879	11 Q9QVS2	Q9qvs2 mus musculu
11	120	3.8	738	5 Q9V4U3	Q9v4u3 drosophila
12	120	3.8	870	5 Q9N4T8	Q9n4t8 caenorhabdi
13	120	3.8	872	4 Q9H3N6	Q9h3n6 homo sapien
14	117	3.7	264	5 Q9WIP3	Q9wip3 drosophila
15	116	3.6	589	4 Q9UGS9	Q9ugs9 homo sapien
16	116	3.6	877	4 Q9UGT0	Q9ugt0 homo sapien
17	115	3.6	1199	11 Q9EPV6	Q9epv6 mus musculu
18	113.5	3.5	940	13 Q9V3635	Q9v3635 fugu rubrip
19	111.5	3.5	362	13 Q9PSY1	Q9psy1 carassius a

20	111.5	3.5	408	13	Q93558	O93558 carassius a
21	110.5	3.5	3507	5	Q23587	Q23587 caenorhabdi
22	110	3.4	540	5	Q9YIA5	Q9yla5 lymnaea sta
23	107.5	3.4	877	13	Q9PW88	Q9pw88 carassius a
24	107.5	3.4	2437	11	Q63725	Q63725 rattus norv
25	107	3.3	383	11	Q70534	Q70534 rattus norv
26	107	3.3	848	13	Q93553	Q93553 carassius a
27	107	3.3	1956	4	Q9Y5Y9	Q9y5y9 homo sapien
28	106	3.3	350	13	Q93556	Q93556 carassius a
29	106	3.3	551	13	Q9PWQ0	Q9pwq0 fugu rubrip
30	105	3.3	383	11	Q62779	Q62779 rattus norv
31	104.5	3.3	3312	4	Q9NY07	Q9ny07 homo sapien
32	102.5	3.2	779	11	Q35269	Q35269 rattus norv
33	102.5	3.2	797	4	Q95288	Q95288 homo sapien
34	102.5	3.2	3680	5	Q9VR08	Q9vr08 drosophila
35	101.5	3.2	864	13	Q73637	Q73637 fugu rubrip
36	101.5	3.2	934	13	Q9DER4	Q9der4 gallus gall
37	101	3.2	1152	11	Q70375	Q70375 mus musculu
38	100.5	3.1	2160	3	Q13328	Q13328 magnaporthe
39	100.5	3.1	2160	3	Q13488	Q13488 magnaporthe
40	100.5	3.1	3313	11	O88278	O88278 rattus norv
41	100	3.1	1062	11	O60789	O60789 mus musculu
42	100	3.1	1993	5	P90670	P90670 aplysia cal
43	99.5	3.1	250	13	Q73645	Q73645 fugu rubrip
44	99.5	3.1	2471	4	Q04721	Q04721 homo sapien
45	99.5	3.1	2471	4	Q9H240	Q9h240 homo sapien

ALIGNMENTS

RESULT 1

Q9VR40	ID	Q9VR40	PRELIMINARY;	PRT;	669 AA.
AC	Q9VR40;				
DT	01-MAY-2000 (TREMBLrel. 13, Created)				
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)				
DE	CG11923 PROTEIN.				
GN	CG11923.				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPRAIN-BERKELEY;				
RA	MEDLINE=20196006; PubMed=10731132;				
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,				
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,				
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,				
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,				
RA	Anderson R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,				
RA	Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,				
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,				
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,				
RA	Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,				
RA	Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,				
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,				
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,				
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,				
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,				
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,				
RA	Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,				
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,				
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houch J.,				
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,				
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,				
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,				
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,				
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,				
RA	Merkulov G., Milshina N.V., Mobarry J., Morris J., Moshrefi A.,				


```
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE GLU-RA PROTEIN.
GN GLU-RA OR CG11144.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fodor C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
Science 287:2185-2195(2000).
EMBL: AE003846; AAF59402.1; -.
FlyBase: FBgn0019985; Glu-RA.
DR InterPro: IPR000337; -.
DR InterPro: IPR001828; -.
DR Pfam: PF00003; 7tm_3; 1.
DR PRINTS: PR00248; GPCRMR.
DR PROSITE: PS00979; G_PROTEIN_RECF_F3_1; 1.
DR PROSITE: PS00980; G_PROTEIN_RECF_F3_2; 1.
DR PROSITE: PS00981; G_PROTEIN_RECF_F3_3; 1.
DR PROSITE: PS00982; G_PROTEIN_RECF_F3_4; 1.
SO SEQUENCE 976 AA; 108485 MW; 43A0E1F918EDACC4 CRC64;

Query Match
Best Local Similarity 5.08; Score 161.5; DB 5; Length 976;
Matches 101; Conservative 20.18; Pred. No. 6.2e-05;
Indels 143; Gaps 21;

Qy 145 SREONLQDDLWYQALV-----SLLEGPSISRAAITSTDS-LSAPAPQVFOATRES 199
Db 295 AKRNLSPQFHWIADSGWKQKLLGLELDAEAGNITVELSEIADFDRIYMQLTPTN 354
```

Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
 "The genome sequence of *Drosophila melanogaster*."
 Science 287:2185-2195(2000).
 EMBL: AE003635; AAF53175.1; -
 FlyBase: FBgn0032418; CG17215.
 InterPro: IPR000337; -
 InterPro: IPR000834; -
 InterPro: IPR002524; -
 Pfam: PF01545; Cation_efflux; 1.
 PROSITE: PS00133; CARBOXYPEPT_N2; UNKNOWN_1.
 PROSITE: 1677 AA; 188028 MW; F50A8D282A8E3B6E CRC64.
 SQ SEQUENCE

Query Match 4.7%; Score 149; DB 5; Length 1677;
Best Local Similarity 14.3%; Pred. No. 0.0014;
Matches 58; Conservative 37; Mismatches 83; Indels 228; Gaps

Qy	265	WSPPYLECENGSYKPGWLVTLSAI-----YGLQNLPVPEFRGVMKVNDINLQKYDIDQCS	319
Dd	99	WTVPYFSCQSRR---WLVSYSIAIPIGRHGL-----RGFISIDIVSTLRVNQCE	146
Qy	320	SDGW-----PSGTHKHLNASE	336
Dd	147	AEPYPFGSRQKMQLOTHNLGARRSLFLSGRMGAIDESTINDLOAFSSHKHCH--RTS	204
Qy	337	CMPITKGL-----	343
Dd	205	MVPVKRLNRKIKEPVSPKLSANQVTPTRRIKRINTPVAVDPSPSVVARARTSAALK	264
Qy	344	-----	343
Dd	265	SFINKYRREHPGKLRAEVVSIWRKMSVEEKQAOFRTIGTETHVFHEBEVPVPSIND	324
Dd	344	-----GFVLGAYECICRAGFYHPGVLPNNFRR	371
Dd	325	VIDIVECDYRQPSEAFTPTVTGKKLLTTLTGSSSWTRGSYQCLCRGGFY-----SL	374
Qy	372	RGPD-----QHISGSTKDVSEEAUYLCPLCRECCPCADDSPCFOEDKYLRLA	420
Dd	375	RHPDGNGTILMELAWQEODNISNYSEVFKCLCAPCDDCTCTGPECLANYHWPFRI	434
Qy	421	ISFGGLCMLLDFVSMVLVVYHFRAKSIRASGLILETILFGLSLLLY	466
Dd	435	LTISICACGTFVIAGLYLFHRHRVRVFKVASPFIIMTLIGCAIMY	480

RESULT	5	
073636		
ID	073636	PRELIMINARY;
AC		PRT; 868 AA.
DT	073636;	
DT	01-AUG-1998	(TREMBLrel. 07, Created)
DT	01-AUG-1998	(TREMBLrel. 07, Last sequence update)
DT	01-JUN-2000	(TREMBLrel. 14, Last annotation update)
DE		PHEROMONE RECEPTOR.
GN	CA02.1.	
OS	Fugu rubripes (Japanese pufferfish)	(Takifugu rubripes).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;	
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;	
OC	Tetraodontidae; Takifugu.	

NCBI_TaxID=31033;	
[1]	
SEQUENCE FROM N.A.	
MEDLINE=98226788; PubMed=9560249;	
Naïto T., Saito Y., Yamamoto J., Nozaki Y., Tomura K., Hazama M.,	
Rakanishi S., Brenner S.;	
Putative pheromone receptors related to the Ca2+-sensing receptor in	
Fugu.;	
Proc. Natl. Acad. Sci. U.S.A. 95: 5178-5181(1998).	
EMBL: AB008858; BAA26123.1; -.	
InterPro: IPR000337; -.	
InterPro: IPR001828; -.	
InterPro: IPR002052; -.	
Pfam: PF00003; 7tm_3; 1.	
Pfam: PF01094; ANF_receptor; 1.	
PRINTS: PR00248; GPCRMR.	
PROSITE: PS00981; G_PROTEIN_RECP_F3_3; 1.	
PROSITE: PS00092; N6_MTASE; UNKNOWN_1.	
Pheromone.	
868 AA; 95531 MW; 3FC66EB1E9972E01 CRG64;	
SEQUENCE	

Query Match 4.1%; Score 132.5; DB 13; Length 868;
Best Local Similarity 19.7%; Pred. NO. 0.015;
Matches 113; Conservative 74; Mismatches 211; Indels 177; Gaps

Qy	119	HPSLHRALDITLTHATNFMVLMQSNKRSREQLQD-----DLDWTQALVWMSLEGEPSI	171
Db	279	HSRIKRVADYIRRSATAVVVAETASTEMMILLUELSELSHEPSPQWIGSESWVT---DPDL	335
Qy	172	SRAAITSTDSLSPAPQVFLQATREESRILLDGLSSAPHLANATLETFWPHGLRRKR	231
Db	336	LR--FSCACGTIGATQRSVIPGLRD-----FLDLSPS--KVASSPVLTEF-----WE	380
Qy	232	PHLHRRPNQGRGLGHSWRKKDGLGDKSHFKWSPPY-----LECENSGYKPGWL	283
Db	381	DSFNCR-----LGKGERMCD--GSEIDMTLQSPYTDTSRLTNMVKAVYAIHA	429
Qy	284	-----TLSSAIYGLQNLVPEFRGVKMKVDINLOKVYDIDQSSDGGWFSGTHKCH	331
Db	430	IHNACVODTNATRCSKFTTINPK-----KVLTLKTNFNSO-----	466
Qy	332	LNNSECMPIRGLGFVLGAYECI--CRAGFYHPCVLPVNNF-----	369
Db	467	--NGYAVSPDANGDPVASYELVNMKSGSGSIEVPVGYDYDASLPBGQEFRI	524
Qy	370	-RRGPDQHTSGS-----TKDVSEE-----AYVCLPCREG-----CPCADD	405
Db	525	GRKQVPVSGSDSCPQGRKVLQKGRKPICYDCVQCPGEIGSEINVTDSPECIC	584
Qy	406	---SPCFVQEDKYLR-----LATISFQGLCMLLDFVSMVLVYVYHFRKAKS	454
Db	585	PERNACEPKPVFELSNEVLGIIILAVFSGGACCLAV--ITAAVFFHHRTPSIVR	642
Qy	455	LETILFSGLLLYFPVWILYPEPSTFCRILLRWARLGFATVGTVTLKHURLVK	514
Db	643	SFULLFSLTLCFCLSLTFIAPSLHSCMLRHRTAFGTFVFLCISCVLGKTVV	701
Qy	515	AQRIPYMTGGRVM-----RMLAV-----LLLWVFWFLIGWTSVCONLEK	560
Db	702	-----LPGSNVMKWFPGPQORMTVVTFTSIQVLICIVMLVWNPFPV--RNL	746
Qy	561	QGKTSOHLIFNMCLIDRWDMYMTAVGMWSLVSYDGL	595
Db	747	--TTYKERIILEALG-----SSVGFWAVLGYIGL	774
RESULT	6		
Q9VPS7	Q9VPS7	PRELIMINARY;	PRT: 1305 AA.
ID	Q9VPS7		
AC	Q9VPS7		
DT	01-MAY-2000 (TReMBLrel. 13, Created)		
DT	01-MAR-2001 (TReMBLrel. 16, Last sequence update)		

RESULT	6
Q9VPS7	
ID	Q9VPS
AC	Q9VPS
DT	01-MA
DT	01-MA

[illegible][illegible]


```
Matches 91; Conservative 65; Mismatches 185; Indels 159; Gaps 21;
QY 138 VMLGNSKRE-----QNIQDLWDYALV-----SLLEGESISRRAITFTDLSAPAP 188
Db 274 LFMRSDDSLIAAASRVNASFTVWASDGMGAOESIVKGSBHVAYGATLELAS----- 327
QY 189 QVFLQATRESRILLQDLSSAPHLANATLETETFWHGLRRKWRPHILHRRGNQPRGLGH 248
Db 328 ----HPVRQFDR-YFQSLNPNYNNH-----RNPWFDRF----- 354
QY 249 SWRRK--DGLGCKSHKSPSPYCELENGSYKP-----GWLVTLSAAYGLQPNLVP 298
Db 355 -WEQKFOCSLQNRHRCIDKHLAIDSSNYEQESKIMFVNVNAVYAMAHLRKWRQTLCP 413
QY 299 EFRGMKVDINLQKVDIDQCSSDQWGFSTHCKHLNNSECMPKIGLGVFLGAYECICKAGF 358
Db 414 ---NTTKCDAMKILDKKLYKDYLLKINFAPEN-----PNKGA-----DSTVRFDT 458
QY 359 YHGVLPVNNRRRGGPDQHIHSGS-----TKDVSSE-- 388
Db 459 YGDGMGRYNFEN-----FOHIGGKYSYLVKGHWAETLYLDVDSIHWSRNSVPTSCSDPCA 514
QY 389 -----AYVCLPCRE-----GCPCFC-----ADDSPCFVQEDKYL-- 417
Db 515 PNEMKNQPGDVCCWICIPCEPYEYLYVDEFTCMCGGQWPTADLSGCYNLPEDYIRWED 574
QY 418 -----LATISFOGLCMLDFVSMVYVYHFRKAKSIRASGLILLETILFGSLLLYFPV 469
Db 575 AWATGPTVIACLGPMCTCIVI-----TVFIKHNTPLVKASGRELCYILLFGVSLSCMT 629
QY 470 VILFEFSTFCILLRWLRLLGFATVGTVTLKLRVLKVP--LSRVAQRIPYMT--GGRV 526
Db 630 FFFAKSPVICALRRRLGLGTSFATCYSALITKTNCTARIFDGVKNGAQRKPFISPSQV 689
QY 527 MRMLAVIL---LVVFWFLI 542
Db 690 FICGLLILQIVMVSVWDIL 709

RESULT 11
Q9V4U3 PRELIMINARY; PRT: 738 AA.
ID Q9V4U3
AC Q9V4U3
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CG8692 PROTEIN.
GN CG8692
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
RP STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
```

```
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houch J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel M.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Ie Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirski R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR FlyBase: FB0033298; AAF59082.1; -.
DR EMBL: AE003837; AAF59082.1; -.
DR InterPro: IPR000337; -.
DR InterPro: IPR001828; -.
DR Pfam: PF000003; 7tm_3; 1.
DR Pfam: PF01094; ANF_receptor; 1.
DR PRINTS: PR00248; GPCRMR.
DR PROSITE: PS00981; G_PROTEIN_RECEP_F3_3; 1.
SQ SEQUENCE 738 AA; 83179 MW; 0E620337B059EAC CRC64;
```

Query Match 3.8%; Score 120; DB 5; Length 738;

Best Local Similarity 18.8%; Pred. No. 0.13;

Matches 106; Conservative 63; Mismatches 198; Indels 196; Gaps 26;

```
QY 70 LAQKLAEEVPMDVASYLYTGDS-----HQLKRANCSGRYELAGLPKWPALASAPSLH 123
Db 62 IVQKLLTK-PRAGAIIFGSDQEVQRQVMRAVRANATGSFWSIGSDG-WSA----- 110
QY 124 RALDTLTHATNFLNVLQSNKSRQNLQDLDWYQALVWSLLEGEPISRRAITFTDSL 183
Db 111 -----RNLVSD-----DYEPEVEG-----TLVQVQ 131
QY 184 SAPA---POVFLOATRESRILLQDLSSAPHLANATLETETFWHGLRRKWRPHLHRRGN 240
Db 132 ANPYRGFEYFLSLTVENQ-----RNPWFVEF---WEDHFCQRYPG 170
QY 241 QGPRGLGHSWRKDKGLGDKSHFKWSPY-----LECNGSYKPGWLVTLSA 288
Db 171 -----STSTPYNNYTKOCTTKERLSRQNTDFE-DQLQFVSDA 206
QY 289 I-----YGLQ-----PNLVPFRCGMKVDI---NLQKVDIDQSSDGN-FSGT----- 327
Db 207 VMAFAYALRDMIRDLCCGGSPCLCEAMKPTKGADLLKYLRKVEFEGSLGDEFDGDGDP 266
QY 328 ---HKCHLNNSECMPIKGLGVFLGAYECICKAGFYHPGVLPWN-----NFRRRGPDQHS- 379
Db 267 ARYNIHFQSQQA-----GOYHWV-KVGETEGELRLNMTVEFKRLSPKPPESV 315
QY 380 -----GSTKDVSEAYVCLPC-----REGCPFC-----ADDSPCFVQE 412
Db 316 CSLPCLVGOAKKYVEGSCCHWCNCTTYQIRHPDDETHCKLCKLGLTLPDAHKKQYCRIP 375
QY 413 DKYLR-----LATISFOGLCMLDFVSMVYVYHFRKAKSIRASGLILLETILFGSLLLY 466
Db 376 EYLRPESAWAGAMAFSATGILVTLFVNGFVVRHNDPIVRASRELSYILLAGIFMYC 435
QY 467 FPMVILYEPSTFCILLRWLRLLGFATVGTVTLKLRVLKVELS--RTAQRIDPYMTGG 524
Db 436 GVTFLVLPKPTNIVCAIORFGVGCTVVYAAALLTKTNRIARFRAGKQSAKRPFSISK 495
```

```
QY 525 RVMRLAV-----ILLVWFLLI 542
DE : : |
GN : : |
DB 496 SOLVICACLSVQILLINGVMV 518

RESULT 12
Q9N4T8 PRELIMINARY; PRT: 870 AA.
ID Q9N4T8
AC Q9N4T8;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE Y4C6A.2 PROTEIN.
GN Y4C6A.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
NCBI_TaxID=6239;
[1]
SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Layman D., Graves T., Yoakum M.;
RT "The sequence of C. elegans cosmid Y4C6A.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006791; AAF60738.1; -
DR InterPro; IPR000337; -
DR InterPro; IPR001828; -
DR Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 2.
DR PRINTS; PR00248; GPCRMR.
SQ SEQUENCE 870 AA; 97727 MW; 2FCE2F871D6B1EF9 CRC64;

Query Match 3.8%; Score 120; DB 5; Length 870;
Best Local Similarity 22.7%; Pred. No. 0.16;
Matches 49; Conservative 35; Mismatches 102; Indels 30; Gaps 5;

QY 390 YVCLPCR-----EGCP-----FCADDSPCFQEDKYLRLLAIIISFQGLCM 428
DB 566 WACIPCDTSTSIHNETSCCEACVGMVDPRTLHFCVPIPPVSMQWDITWSLIPAAFTSLGI 625

QY 429 LIDFVSMVVYHFRKASTRASGLILLETILFGSLLYFPVILYFEPSTFCILLRWAR 488
DB 626 ASTIFVSVFLKSNTPVIMASGRELICYMMSGIGMCTLTFFLVSQPTVITCSMTIRLM 685

QY 489 LIGFATVYGTVTLKLRHLVLFSLRTAQRIPYMTG---GRVMRLAVILLVVF-WFLIG 543
DB 686 GLSMNAIYAAITKTNRLARVFPDSQAPRPTKPAQVIGCMGVISVLIQTFTVWILFD 745

QY 544 WTSVVCQNLEKQISLIGQCKTSDH-----LIFNMCLI 575
DB 746 PPGTMIVFPTRTEAVLTCKATTSHLLISLLYNILLI 781

RESULT 13
Q9H3N6 PRELIMINARY; PRT: 872 AA.
ID Q9H3N6
AC Q9H3N6;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
```

```
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR TYPE 2.
GN HMG1UR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Yasuyuki F., Akiko J.;
RT "Structure and polymorphisms of the human metabotropic glutamate
RT receptor type 2 (hmg1ur2) gene: Analysis of association with
RT schizophrenia.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045011; BAB19817.1; -
KW Receptor.
SQ SEQUENCE 872 AA; 95567 MW; 801976D034AA8100 CRC64;

Query Match 3.8%; Score 120; DB 4; Length 872;
Best Local Similarity 26.2%; Pred. No. 0.16;
Matches 49; Conservative 28; Mismatches 70; Indels 40; Gaps 9;

QY 390 YVCLPCR-----CADD-----SFCVQEDKYLR-----LAIIIS 422
DB 520 WLCIPQC---PYEYRLDEFTCADCGLYWPNASLTGCFELPQEIYRWGDAWAVGPVTIAC 576

QY 423 FQGLCMILDFVSMVYHFRKASIRASGLILLETILFGSLLYFPVILYFEPSTFCRI 482
DB 577 LGALATL--FVLGVFVRH-NATPVVRASGRELICYLLGGVFLCYCMTFFIAKPSTAVCT 633

QY 483 LLRWARLLGFATVYGTVTLKLRHLVLF--LSRTAQRIPYMT--GGRVMRLAVI----LL 535
DB 634 LRRLLGTAFSVCYCSALLTKTNRIARIFGARGAQRPRFISPASQVAICLALISQLLI 693

QY 536 VVFWFLI 542
DB 694 VVAVLVV 700

RESULT 14
Q9W1P3 PRELIMINARY; PRT: 264 AA.
ID Q9W1P3
AC Q9W1P3;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE CG18678 PROTEIN.
GN CG18678.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
```

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
FT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RA EMBL; AE003461; AAF47013.1; -.
RA FlyBase; FBgn0040664; CG18678.
RA InterPro; IPR000337; -.
DR Pfam; PF00003; 7tm_3; 1.
DR PRINTS; PR00248; GPCRMR.
SQ SEQUENCE 264 AA; 29532 MW; CAC5623C52703642 CRC64;

Query Match 3.7%; Score 117; DB 5; Length 264;

Best Local Similarity 26.1%; Pred. No. 0.064;

Matches 29; Conservative 23; Mismatches 47; Indels 12; Gaps 3;

QY 472 LYEPSPFRCLLWARLLGFATVGVVTLKLHRLVKLVFLSRTAQRIPYMTGGVRMRLA 531
DB 1 MIPNLTCTARLWLREIGFSLTYGALMLKTRISVIFVRSAKAK-ITDAALLRLG 59
QY 532 VILLVFWFLIGWTSVQCNLEKQISL--ICQKGTSDHLFNMCLIDRWYD 580
DB 60 IICGAI-----GTCLLVRLVSPDPVVVGRVTTADDLKAFCKTDWDY 101

RESULT 15

Q9UGS9
ID Q9UGS9 PRELIMINARY; PRT; 589 AA.
AC Q9UGS9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DT DJ69B13.1 (GLUTAMATE RECEPTOR, METABOTROPIC 1 (BETA ISOFORM))
(FRAGMENT).
GN GRM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bates K.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035698; CAB65992.1; -.
DR InterPro; IPR000337; -.
DR InterPro; IPR001828; -.
DR Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMR.
DR PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1.
DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
KW Receptor.
FT NON_TER 1 1
SQ SEQUENCE 589 AA; 66304 MW; BEE9089D87293243 CRC64;

Query Match 3.6%; Score 116; DB 4; Length 589;

Best Local Similarity 22.1%; Pred. No. 0.22;

Matches 49; Conservative 34; Mismatches 95; Indels 44; Gaps 6;

QY 357 GFYHPGVLPVNNFRRGPDQHIISGSTDVDSEE-----AYVCLPCREG- 398
DB 181 GTWHEGVNLNDDYKIQ---MNKSGVRSVCSECLKGQIKVIRKGEVSCCWCICKAKENE 237
QY 399 -----CPFC-----ADDSPCFVQEDKYLR-----LAIISFQGLCMLLDFVSMLV 438
DB 238 YVQDEFTCKACDLGWPNADLTGCEPIPVRYLEWSNIESIIAIFSCGLIYLVTLFVLIF 297
QY 439 YHFRKAKSIRASGLILLETILFGLSLLYFPVVILYFPFPRCILLRWALLGFATVYGT 498
DB 298 VLYRDTPVWRSSSRELCYIILAGIFLGVCPFTLIAKPTTTSYQLRLLVGLSSAMCYSA 357
QY 499 VTLKLRVLKVFL---SRTAQRIPYMTGGVRMRLAVILLVW 537
DB 358 LVYTKNRIARILAGSKKKIKCTRKPRFMSAWAQVITIASILISV 399

Search completed: November 1, 2001, 10:19:24

Job time: 352 sec

OM of: US-09-775-181-4 to: EST.* out_format : pfs

Date: Nov 1, 2001 10:59 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+*_p2n.model -DEV=xlh
-Q/cgn2/_USPTO.spool/US09775181/runat_30102001_161056_13799/app_query.fasta_1.1955
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -XGAPOP=6.000
-XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=bgsum62 -TRANS=human40.cdi
-ALIGN=15 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09775181_@CGL1_1.7084 -NCPU=6
-ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPPY
-WAIT -THREADS=1

Search information block:

Query: US-09-775-181-4

Query length: 599

Database: EST.*

Database sequences: 10228115

Database length: 431459454

Search time (sec): 2318.790000

score_list:

Sequence	Strd	Orig	zScore	Escore	Len	Documentation
gb_gss3:CNS02NOT	-	398.00	646.87	6.5e-27	632	AL205670 Tetraodon nigroviridis
gb_gss3:CNS022LV	-	397.50	640.77	1.4e-26	961	AL178348 Tetraodon nigroviridis
gb_gss4:CNS03E8	-	323.50	516.19	1.2e-19	1030	AL240569 Tetraodon nigroviridis
gb_gss4:CNS03H19	-	293.00	467.22	6.1e-17	877	AL243702 Tetraodon nigroviridis
gb_gss4:CNS03235	-	282.00	456.34	2.7e-16	483	AL290325 RPCI-24-187D3.TJ RPCI-
gb_gss4:CNS0304C	-	235.50	369.73	1.8e-11	978	AL252885 Tetraodon nigroviridis
gb_gss3:CNS0317B	-	229.00	361.28	9.9e-11	807	AL223184 Tetraodon nigroviridis
gb_gss3:CNS0203P	-	228.00	356.38	9.9e-11	1043	AL175678 Tetraodon nigroviridis
gb_gss3:CNS02FTA	-	227.50	358.54	7.5e-11	822	AL195463 Tetraodon nigroviridis
gb_gss3:CNS020UA	-	226.00	354.12	1.3e-10	957	AL207163 Tetraodon nigroviridis
gb_gss5:CNS05CMB	-	217.00	337.37	1.1e-09	1096	AL331292 Tetraodon nigroviridis
gb_gss4:CNS03YL7	-	212.50	332.49	2.1e-09	888	AL2266452 Tetraodon nigroviridis
gb_gss4:CNS035FK	-	210.00	329.01	3.3e-09	840	AL228665 Tetraodon nigroviridis
gb_gss30:AZ628814	+	201.00	325.76	5.0e-09	329	AZ628814 IM0481D17F Mouse 10kb
gb_est27:AJ280627	+	200.00	318.38	1.3e-08	518	AJ280627 A3A-AAR-F-11-F Anophe
gb_gss4:CNS03EM7	+	194.00	312.28	2.8e-08	379	AL240568 Tetraodon nigroviridis
gb_gss4:CNS03X7G	+	188.50	290.73	4.5e-07	1012	AL265957 Tetraodon nigroviridis
gb_est7:AA439256	+	169.50	284.97	1.2e-05	628	AA439256 LD13768.5prime LD Dros
gb_est29:AU004575	+	159.50	252.55	6.0e-05	446	AU004575 AU004575 Bombyx mori p
gb_gss4:CNS04BT8	+	159.00	242.27	0.0002	945	AL283589 Tetraodon nigroviridis
gb_gss33:AZ2823042	-	156.00	244.82	0.0002	518	AZ823042 2M0096N15R Mouse 10kb
gb_gss2:CNS01307	-	155.00	235.51	0.0005	951	AL103321 brosofilia melanogaste
gb_est10:AA697422	+	153.00	236.09	0.0005	696	AA697422 HU02444.5prime HL Dros
gb_gss3:CNS01XPU	+	151.00	228.55	0.0013	972	AL172011 Tetraodon nigroviridis
gb_hlc:AK002958	+	127.50	181.97	0.5122	1736	AK002958 Mus musculus adult ma
gb_gss5:CNS050V4	-	123.00	180.19	0.6430	1099	AL316057 Tetraodon nigroviridis
gb_gss5:CNS05DM1	-	121.00	176.90	0.9814	1095	AL332938 Tetraodon nigroviridis
gb_est102:BG540839	-	120.00	173.48	1.52	1258	BG540839 602570768F1 NIH_MGC_7
gb_est24:AZ294454	-	117.00	175.75	1.14	705	AZ294454 RPCI-23-132H13.TV RPCI-
gb_est49:AW581952	-	115.50	177.14	0.9511	517	AW581952 MR4-ST0118-040100-034-
gb_gss3:CNS020HY	+	115.00	169.73	2.46	872	AL206719 Tetraodon nigroviridis
gb_est99:BG343241	+	114.50	166.89	3.54	1023	BG343241 HVSM90005D05f Hordeu
gb_gss3:CNS02030	+	114.50	166.58	3.68	1048	AL175677 Tetraodon nigroviridis
gb_hlc:AA021164	+	114.50	156.73	13.04	2294	AA021164 Mus musculus 16 days
gb_est9:AK018743	+	113.50	172.05	1.83	594	AK018743 SMOV3MCA1057SK Onchoce
gb_est83:BF123539	+	113.50	166.36	3.79	934	BF123539 601759855F1 NCI CGAP_M
gb_est84:BF166785	-	111.00	162.22	6.45	931	BF166785 601775106F1 NCI CGAP_I
gb_est97:BG177951	+	111.00	158.33	10.61	1268	BG177951 602327879F1 NIH_MGC_9
gb_est81:BG962924	+	111.00	155.29	15.68	1615	BG962924 601656470R1 NIH_MGC_6
gb_est84:BF966904	-	110.50	160.56	7.98	994	BF966904 602286594T1 NIH_MGC_95

gb_est72:BE299868 + 110.00 162.04 5.60 827 ! BE299868 600944667F1 NIH_MGC
gb_est95:BG029835 - 110.00 161.50 7.07 863 ! BG029835 602296569F1 NIH_MGC
gb_gss28:AZ540053 + 110.00 160.64 7.89 924 ! AZ540053 ENTELO9TF Entamoeba
gb_est86:BF347266 - 109.50 159.90 8.68 917 ! BF347266 602021013F1 NCI CGA
gb_gss3:CNS02AUZ + 109.50 158.45 10.45 1029 ! AL189044 Tetraodon nigrovir

seq_name: gb_gss3:CNS02NOT

seq_documentation_block:

LOCUS CNS02NOT 632 bp DNA 14-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
151N10 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION AL205670.1 GI:7864489

VERSION GSS; genome survey sequence.

KEYWORDS Tetraodon nigroviridis.

SOURCE Tetraodon nigroviridis

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 632)

AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 632)

AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.

TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 632)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES Location/Qualifiers

source 1..632

organism="Tetraodon nigroviridis"

db_xref="taxon:99883"

/clone_lib="G"

/note="Genoscope sequence ID : COAG151DG05LP1-end : T7"

BASE COUNT 99 a 206 c 219 g 99 t 9 others

ORIGIN ..

alignment_scores:

Quality: 398.00 Length: 201

Ratio: 2.707 Gaps: 7

Percent Similarity: 73.134 Percent Identity: 46.766

alignment_block:

US-09-775-181-4 x CNS02NOT/rev ..

Align seg 1/1 to reverse of: CNS02NOT from: 1 to: 632

104 GluLeuAlaGlyLeuProGlyLysrpproAlaLeuAlaSerAlaHisPr 120

||||| :||| |||

630 GAGCTCCCTCTCTCTGGGGAAGKTCACG.....GCCGCCGCCGCG 587

120 oSerLeuHisArgAla.LeuAspThrLeuThrHisAlaThrAsnPhelLeu 136

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

586 CTCGGTTACGCGCGGCGGAGCGGCTGCTGCACGCGCCCAACTCTTC 537


```
67 yThrileuAlaGlnLysLeuAlaGluValPrometaspValAlas 84
188 .....CTGG 191
84 erTyrlLeuTyrrThrGlyAspSerHisGlnLeuLysArgAlaAsnCysSer 100
.....CTGG 100
192 CCTTTCTCTACTCTGAGAGTGTCAACGGCTGTCAAGGAGCAAACTGCAGT 241
101 GlyArgTyrrGluLeuAlaGlyLeuProGlyLysTrpProAlaLeuAlase 117
.....CTGG 117
242 GAGAAATATGAAGTCGTGGGGCGGAAGCAAA.....GCCGG 279
117 rAlaHisProSerLeuHisArgAlaLeuAspThrLeuThrHisAlaThrA 134
.....CTGG 134
280 GTGCCCCCAGCTTCACAGAGAGCGGGCACCTTGGCCAGGCTGCCA 329
134 snPheLeuAsnValMetLeuGlnSerAsnLysSerArgGluGlnAsnLeu 150
.....CTGG 150
330 ACTTTCTCAACATGCTCTACAGCCACAGCATCCGGGAGTCGAGCGTG 379
151 GlnAspLeuAspTrpTyrrGlnAlaLeuValTrpSerLeuLeuGluG1 167
.....CTGG 167
380 GAGGAGGAGCGTGGAGTGTATCAGCGCTGTTCGACGTGTGGCAGAGGG 429
167 yGluProSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeu 184
.....CTGG 184
430 AGACCCGAAGCCTACAGGGCTCTGCTGACCTTTAAC.....CTGG 466
184 erAlaProAlaPro 188
467 .....CTGGCACC 475
```

seq_name: gb_gss4:CNS0304C

```
seq_documentation_block: 978 bp DNA GSS 17-MAY-2000
LOCUS CNS0304C Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 041G06 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL252985
VERSION AL252985.1 GI:7973897
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 978)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billaut,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 978)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 978)
Genoscope.
Direct Submission
TITLE Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL This sequence is a single read and was generated as part of a large
REFERENCE scale clone-end sequencing project of the Tetraodon nigroviridis
AUTHORS genome. For more information, please take a look at
JOURNAL http://www.genoscope.cns.fr/Tetraodon.
COMMENT
FEATURES
source Location/Qualifiers
1..978
/organism="Tetraodon nigroviridis"
```

```
/db_xref="taxon:99883"
/clone="041G06"
/note="Genoscope sequence ID : COBG041BD03LP1-end : T7"
BASE COUNT 176 a 267 c 344 g 188 t 3 others
ORIGIN
alignment_scores:
Quality: 235.50 Length: 286
Ratio: 1.869 Gaps: 14
Percent Similarity: 44.056 Percent Identity: 26.923
alignment_block:
US-09-775-181-4 x CNS0304C/rev ..
```

```
Align seg 1/1 to reverse of: CNS0304C from: 1 to: 978
182 SerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThrArgGluG1 198
|||||
956 TCCTCCCGCACCCCGCGGCAAGGTG..... 930
198 uSerArgIleLeuLeuGlnAspLeuSerSer.....AlaProHisL 213
|||
929 .CAGACCATCGTCTCCAGGACCTGTCCAAGGCTGGGACACGCTGCATC 881
213 euAlaAsnAlaThrLeuGluThrGluTrpPheHisGlyLeuArgArgLys 229
|||
880 TGCCCGCTCCCGCGGACGACAGCTGGTTCAGGAGT.....TTWAAA 837
230 TrpArgProHisLeuHisArgArgGlyProAsnGlnGlyProArgGlyLe 246
|||
836 TTCCCGCGG.....CCCAATCAG...CCGACAGCCTT 808
246 uGlyHis.....SerT 250
|||
807 GTCCAAACGGGTGCTCTTAAGACCTCAGCACCTGGACACGCCAAGT 758
250 rpArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrp... 265
|||
757 GGGGGCGGGGCGACAGCTACGTGACCAACCGCGGCGCGCTGGGCC 708
266 Ser.ProProTyrrLeuGluCysGluAsnGlySerTyrrLysProGlyTrpL 282
|||
707 AGCGCCCGCGCTCGACGTGCCAGACGGCGCTTCGTGCTGCTGCTGGA 658
282 euValThrLeuSerAlaIleTyrrGlyLeuGlnProAsnLeuValPro 298
|||
657 CGCTCACCTGTCCATGCCCTTCTAGGGCTCAAGCCCGACCTGACACCG 608
299 GluPhe..... 300
|||||
607 GAGTTTCAGTTCAGGTCTCTCCACACACACACACACACACACACATC 558
300 ..... 300
557 ACACACACGGAAGCACCCCTTCCAAATGATTGAATTTGAATAATGAGGCC 508
300 ..... 300
507 TGACGGAGGCTCAGCGGGCTGTCCCCCTCCNCCCTCCGGTCCGTATGT 458
301 .....ArgIlyValMetLysValAspIleAsnLeuGlnLysValas 314
|||||
457 CTGTCGGTTAGAGCGGTGATCGCGCTGGAGCTCAACATCCAGACATCGA 408
314 pIleaspGlnCys....SerSerAspGlyTrpPheSerGlyThrHisLysC 330
|||||
407 CTGGACCATGTCAGCGGGGGGAGCGGTGTTCCGCCACACCCAGAGT 358
330 yHisLeuAsnAsnSerGluCysMetProIleLys..... 341
|||
357 GC.....AACCGCACCATGAGGTAAAGAAAGCCCTGACCCCGAG 317
```



```

seq_name: gb_gss4:CNS03YL7

seq_documentation_block:
LOCUS      CNS03YL7          888 bp          DNA          18-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
            068H11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION  AL266452
VERSION    AL266452.1   GI:7988226
KEYWORDS   GSS: genome survey sequence.
SOURCE     Tetraodon nigroviridis.
ORGANISM   Tetraodon nigroviridis.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleost
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE  1 (bases 1 to 888)
AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.

```

REFERENCES	Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 888)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .

```

/clone="069H11"
/clone_lib="G"
/note="Genoscope sequence ID : COBG068CD061"
BASE_COUNT 203 a 242 C 296 g 143 t 4 others
ORIGIN

alignment_scores:
    Quality: 212.50      Length: 94
    Ratio: 3.320        Gaps: 3
    Percent Similarity: 68.085    Percent Identity: 46.809

alignment_block:
US-09-775-181-4 x CNS03YL7/rev ..

Align seg 1/1 to reverse of: CNS03YL7 from: 1 to: 888

352 CysIleCysLysAlaGlyPheTyrHisProGlyValLeuProValAsnAs 368
|||
496 TTGGGCTTCCTCTCAGGAGCGGAGGAGGAGCGTCTCCGGGGAGC... 449
|||
368 nPheArgArgGlyProAspGlnHisIleSerGlySerThrLysAspV 385
448 .....CGTCCACGACGACGAG.....GGCTCCTCCAGCGGAC. 416
|||
385 aISerGluGluAlaTyrValCysLeuProCysArgGluGlyCysProPhe 401
|||
415 .....TGCCGTGCCCGCGGAGGGGTGCACCTAC 386
|||
402 CysAlaAspSerProCysPheValGlnGluAspLysTyrLeuArdLe 418

```

```

VERSION      A2628814.1  GI:11751004
KEYWORDS     GSS.
SOURCE       house mouse.
ORGANISM     Mus musculus

REFERENCE    1 (bases 1 to 329)
AUTHORS     Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
            Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
            M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
            and Wright, D., Weiss, R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL      Unpublished (2000)
COMMENT      Contact: Robert B. Weiss
            University of Utah Genome Center
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0481 row: D column: 17
            Seq primer: CGTTGTAACGACGCGCCAGT
            Class: plasmid ends
            High quality sequence stop: 329.
FEATURES     source
            1..329
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC1M0481D17"
                /lab_host="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /note="Vector: PWD42nv; Purified genomic DNA from M.
            Laboratory Mouse DNA Resource
            (http://www.jax.org/resources/documents/dnares/). The DNA
            was hydrodynamically sheared by repeated passage through a
            0.005 inch orifice at constant velocity. The sheared DNA
            was blunt end-repaired with T4 DNA polymerase and T4
            polynucleotide kinase. Adaptor oligonucleotides were
            ligated to the blunt ends in high molar excess. The
            adaptor DNA was purified and size-selected for a 9.5 to
            10.5 kb range using preparative agarose gel
            electrophoresis. Vector DNA was prepared from a derivative
            of PWD42 (g114732114|gblAF129072.1), a copy-number
            inducible derivative of plasmid RL. The vector was ligated
            with adaptors complementary to the insert adaptors and
            purified. The sheared, adaptor mouse DNA was annealed to
            adaptor vector DNA, and transformed into
            chemically-competent E. coli XL10-Gold (Stratagene) cells
            and selected for ampicillin resistance."
BASE COUNT   98 a 55 c 80 g 96 t
ORIGIN

alignment_scores:
    Quality: 201.00      Length: 37
    Ratio: 5.432         Gaps: 0
Percent Similarity: 100.000 Percent Identity: 97.297

alignment_block:
US-09-775-181-4 x A2628814
Align seg 1/1 to: A2628814 from: 1 to: 329
300 PheArgGlyValMetLysValAspLeuAsnLeuGlnLysValAspLeuAs 316
|||||
40 TTTAGGGGTGTGATGAAGTTGATATACCCCTTCAGAAAGTGACATTGA 89

```

```

316 pGlnCysSerSerAspGlyTrpPheSerGlyThrHisLysCysHisLeuA 333
|||||
90 CCAATGTTTCGAGTAGTGCTGTTTTCAGGAAGTCAAAATGCCACCTTA 139

333 snAsnSerGlu 336
|||||
140 ACAACTCAGAG 150

seq_name: gb_est27:AJ280627
seq_documentation_block:
LOCUS      AJ280627      518 bp      mRNA      EST      30-JUN-2000
DEFINITION 4A3A: AAR-F-11-F Anopheles gambiae immune competent 4A3A Anopheles
gambiae cDNA clone 4A3A-AAR-F-11, mRNA sequence.
ACCESSION  AJ280627
VERSION     AJ280627.1 GI:6928508
KEYWORDS    EST.
SOURCE      African malaria mosquito.
ORGANISM    Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea
            ; Anopheles.
REFERENCE   1 (bases 1 to 518)
AUTHORS     Dimopoulos, G., Casavant, T.L., Chang, S., Scheetz, T., Roberts, C.,
            Donohue, M., Schultz, J., Benes, V., Bork, P., Ansoerge, W., Soares, M.B.
            and Kafatos, F.C.
TITLE       Anopheles gambiae pilot gene discovery project: identification of
            mosquito innate immunity genes from expressed sequence tags
            generated from immune-competent cell lines
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6619-6624 (2000)
MEDLINE     20300950
COMMENT     Contact: Dimopoulos G
            Fotis C. Kafatos Laboratory
            European Molecular Biology Laboratory
            Meyerhofstrasse 1, 69117 Heidelberg, Germany.
FEATURES     source
            1..518
                /organism="Anopheles gambiae"
                /strain="4A t/r"
                /db_xref="taxon:7165"
                /clone="4A3A-AAR-F-11"
                /clone_lib="Anopheles gambiae immune competent 4A3A"
                /cell_line="Immune competent 4A3A"
                /lab_host="E. coli DH10B"
                /note="vector: p7T3b-Pac (Pharmacia) with a modified
            polylinker; Site 1: EcoRI; Site 2: NotI; sequenced from
            forward priming site which reads from the 3' end of the
            cDNA. The 4A3A is a directionally cloned and normalized
            cDNA library that was constructed from the 4A3A cell line
            oligo-T primed cDNA according to: Bonaldo, Lennon & Soares
            (1996) : Normalization and Subtraction: Two approaches To
            Facilitate Gene Discovery, Genome Research 6, 791-806."
BASE COUNT   109 a 149 c 129 g 130 t 1 others
ORIGIN

alignment_scores:
    Quality: 200.00      Length: 154
    Ratio: 2.299         Gaps: 6
Percent Similarity: 56.494 Percent Identity: 31.169

alignment_block:
US-09-775-181-4 x AJ280627/rev ..
Align seg 1/1 to reverse of: AJ280627 from: 1 to: 518
262 HisPheLysTrpSerProTyrLeuGluCysGluAsnGlySerTyrly 278
|||||
420 CACGGCTACTGGACGACGCCGCGAGTTCGACTGCAAG.....GGCTACGT 377

278 sProGlyTrpLeuValThrLeuSerSerAlaIleTyrGly.....L 292
|||||
376 GAAGAAGTGGCTCATAAACGTACGCGTCCGCTTCGCGTGGGACAGCC 327

```

```
292 euGlnProAsnLeuValProGluPheArgGlyValMetLysValAspIle 308
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
326 TCAAGCGCAAGCTA.....GAATTCAAGGAGTCGTTGCAGTTTCAATG 283
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
309 AsnLeuGlnLysValAspIleAspGlnCysSerSerAspGlyTyr..... 323
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
282 AACATGCTCCAGCTGGGATATCAACCAATGTCGGGATGATTATTACGAGCC 233
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
324 .....PheSerGlyThrHisLysCysHisLeuAsnAsnSerGluCysM 338
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 GAATGCATTCAAAACACTCACAAATGCGACGAGAGCTTTCATACTGTG 183
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
338 etProIleLysGlyLeuGlyPheValLeuGlyAlaTyrGluCysIleCys 354
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
182 TCCCGATCTCGCGCGGGTACGAAACGGCGGCTATAAGTCGAGTGC 133
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
355 LysAlaGlyPhe.....TyrHisProGlu 362
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
132 CTGCAAGGGTTCGAGTATCCGTACGAGGATCTGATCACCTACTATGACGG 83
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHisIleS 379
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
82 GCAGTGGTCGAGTCCGAGTTCGAGACATCGTGATGATAAG..... 40
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
379 erGlySerThrLysAspValSerGluGluAlaTyrValCysLeuProCys 395
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
39 .....CAATCGCGGTTCGAGACGTTCAAGTGC 13
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
396 ArgGluGlyCys 399
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
12 CGTCTGGCTTGT 1
```